# Message Exchange Games in Strategic Contexts

Nicholas Asher*
*CNRS, Institut de Recherche en Informatique de Toulouse*

Soumya Paul
*Institute de Recherche en Informatique de Toulouse*

Antoine Venant
*Université Paul Sabatier, Toulouse*

## 1 Introduction

Conversations often involve an element of planning and calculation of how best one can achieve one's interests. We are interested in how conversations proceed in a setting in which dialogue agents cannot assume that their interests coincide with those of their interlocutors, and we think this is a promising starting point for a general model of conversation. While there is a large literature in linguistics and in AI on cooperative conversation stemming from Grice (1975), there is little theoretical and formal analysis of conversation in non-cooperative situations. The work of Traum & Allen (1994), where cooperativity is determined only by the social conventions guiding conversation, obligations that do not presuppose speakers to adopt each other's goals, constitutes an important exception. Still, the formal structure of such conversations remains largely unexplored. We propose here a formal theory of message exchange in settings where agents don't necessarily share interests and goals.

In particular, a little explored element in linguistics is the general "shape" of a conversation, its overall structure and the effects of this structure on content. The goals of conversational participants and the context of moves they have already made explain why they make the subsequent discourse moves they do and give a coherence to the conversation as a whole. For conversations where agents share conversational goals and interests, a broadly Gricean answer explored by Grosz & Sidner (1986), Grosz & Kraus (1993) *inter alia* is that the discourse is organized around a problem that it is in the common interest of the participants to resolve; the structure of the conversation reflects the structure of the decision problem, or rather the reasoning of conversational participants to construct a plan that solves the common decision problem.

* Thanks to Julie Hunter, Alex Lascarides, David Beaver, Chris Barker and to the participants of the Rutgers Workshop on Coordination and Content. This work was supported by ERC grant 269427.

Non cooperative conversations, conversations where cooperativity or shared goals and interests cannot be assumed, don't instantiate reasoning about a common decision problem. Consider a debate between two political candidates. Each candidate has a certain number of points she wants to convey to the audience; each wants to promote her own position at the expense of the other's. Strategic conversations are also reactive: to achieve their goals, each participant needs to plan for anticipated responses from the other. To explain "what is going on" in such a conversation, we need to appeal to the participants' discourse goals, which may depend on the goals of the participants' interlocutors. Similar strategic reasoning about what one says is a staple of board room or faculty meetings, bargaining sessions, and even conversations with one's children. These observations indicate that strategic conversations are games, and debates are typically 0 sum games. Typically only one agent can win, though there may also be draws. Such conversations are common.

Grasping the general conversational goals of conversationalists does not suffice, however, to determine the structure of a conversation. Since conversations should be the result of rational inference to the best means for achieving one's conversational goals given one's information about the discourse context, particular *linguistic moves* in a conversation should be related to an overall conversational goal. For cooperative conversations, we need to describe the linguistic reflection of the reasoning about a common decision problem, and this means we need to talk about the way clauses in a text rhetorically relate to each other and how such related clauses can combine to form more complex discourse units bearing rhetorical relations to other discourse constituents in a way that has become familiar not only from Grosz and Sidner but from theories of discourse structure like RST Mann & Thompson (1987) and SDRT Asher & Lascarides (2003). The interaction between goals and particular moves is important for understanding monologue as well, as one can ask what "problem" the author was trying to solve in a particular passage; there is a close correspondence between a coherent text's discourse structure and the text's "goal". We aim to tell a similar story for conversations in not necessarily cooperative settings. Given certain general conversational goals for our conversational participants, we want to track how particular discourse moves detailed in a theory like SDRT takes one dialogue agent towards her conversational goals or thwarts them.

To get a better idea of the structure of conversations in strategic settings, we start from two intuitions. There is a strong intuition that many strategic conversations have a determinate outcome. One dialogue agent can "win" if she can play certain conversational moves; and if she doesn't, she loses. There is also a strong intuition that in many conversations some conversational strategies, and some winning conditions or conversational goals, are more complex, and more difficult to achieve, than others. Understanding and categorizing such winning conditions and their strategies are an important part of understanding the large scale structure of conversations. In addition, they also determine whether a conversational agent has won the strategic conversational game, which is one important communicative effect of the conversation. But how can we measure or compare such strategies? This paper systematizes these intuitions and offers an answer to our question.

To help with these intuitions, here are some examples of conversations with their intuitive winning conditions.

**Example 1.** Suppose a candidate, Candidate A, has a joint interview with another competing candidate, Candidate B, for an academic position. Suppose Candidate A has proved an important theorem and she knows that during the interview if she can mention this, she will have "won" the interview by getting the job over the smarter candidate B, as long as she can mention this fact, no matter at what point of the meeting she says so. This is her winning condition.

**Example 2.** In the following example from Solan & Tiersma (2005), a prosecutor wants Bronston to say whether he had a bank account in Switzerland or not; Bronston does not want to make such an admission. His winning condition is to not answer the question directly, but only to implicate an answer that he doesn't have a bank account. He does not want to commit either to having or to not having a bank account.

(2)
a. Prosecutor: Do you have any bank accounts in Swiss banks, Mr. Bronston?
b. Bronston: No, sir.
c. Prosecutor: Have you ever?
d. Bronston: The company had an account there for about six months, in Zurich.

A non-courtroom variant of (2) is (1). The background is that Janet and Justin are a couple, Justin is the jealous type, and Valentino is Janet's former boyfriend.[1]

(1)
a. Justin: Have you been seeing Valentino this past week?
b. Janet: Valentino has mononucleosis.

1 Thanks to Chris Potts and Matthew Stone for this example.

Janet's response implicates that she hasn't seen Valentino, whereas in fact though Valentino has mononucleosis she has been seeing him him.

**Example 3.** Consider a *voire dire* examination in a medical malpractice suit where the plaintiff lawyer (LP) has as a goal to return repeatedly to the topic about the division of a nerve during a surgery. This goal has a further objective of getting the witness (D) to characterize the surgical operation as incompetent and mishandled. Repeatedly coming back to the topic can wear D down as actually happened in the case we cite.

(3)
a. LP: And also, he put an electrical signal on that nerve, and it was dead. It didn't do anything down in the hand, it didn't make the hand twitch?
b. D: Correct.
c. LP: And we know in addition to that, that Dr. Tzeng tore apart this medial antebrachial cutaneous nerve?
d. D: Correct.
e. LD: Objection.
f. THE COURT: Overruled.
g. D: Correct. There was a division of that nerve. I'm not sure I would say "tore apart" would be the word that I would use.
h. LP: Oh, there you go. You're getting a hint from your lawyer over here, so do you want to retract what you're saying?

The defendant was resisting this line of attack relatively well, but then made an error by agreeing to LP's loaded question.

**Example 4.** During the Dan Quayle-Lloyd Bentsen Vice-Presidential debate of 1988, Quayle was repeatedly questioned about his experience and his qualifications to be President. Quayle attempted to compare his experience to the young John Kennedy's to answer these questions; his winning condition was probably to suggest with this comparison that like Kennedy he was a worthy Presidential candidate. Part of his goal too was to have this comparison pass without criticism (perhaps because he couldn't defend it adequately), and so it was indirect. However, Bentsen made a discourse move that Quayle didn't anticipate.

(4) a. Quayle: ... the question you're asking is, "What kind of qualifications does Dan Quayle have to be president," [...] I have as much experience in the Congress as Jack Kennedy did when he sought the presidency.

b. Bensten: Senator, I served with Jack Kennedy. I knew Jack Kennedy. Jack Kennedy was a friend of mine. Senator, you're no Jack Kennedy.

c. Quayle: That was unfair, sir. Unfair.

d. Bentsen: You brought up Kennedy, I didn't.

Bentsen's surprise move successfully attacked Quayle's strategy to establish a comparison between himself and John Kennedy. Quayle had no effective defense and lost the debate handily.

**Example 5.** Allegedly, the physicist and Nobel laureate Richard Feynmann decided the topics of his next lecture in advance and prepared for it for over 8 hours. However, when he entered the class he would start off with: "So what shall we discuss today?" But he would always have a strategy to steer the conversation to the topics he had prepared for, whatever his students, who always wanted to stump him (and so had opposing interests to Feynmann's), would answer. Feynmann's winning condition was eventually to get to his prepared topic and stick to it for the remainder of the lecture.

Our examples so far have described or been excerpted from actual finite conversations that are relatively circumscribed. But conversations can occur over a much longer period, say over an entire Presidential campaign as in our next example. Nevertheless, they are still *linguistic conversations*.

**Example 6.** Recall President Clinton's adage "it's the economy stupid." What Clinton meant is that he should keep the conversation focussed on questions concerning the economy in the extended debate between his Democratic team and the opposing Republican one during the 1992 Presidential campaign. As long as Clinton was able to bring the debate repeatedly back to a discussion of the economy, he achieved his winning condition.

We claim the following are important features of strategic conversations (and perhaps of conversations generally).
(I) People have conversations for purposes. Their conversations are successful when they achieve those objectives. Crucially, some of these objectives involve commitments to contents by other conversational participants. In all conversations, including those where one person's gain from the conversation

is another person's loss, the interlocutors' contributions force them to commit to certain contents, which are the conventional meanings and implicatures of their utterances in the context.
(II) In principle, conversational players have no limits on the length of their intervention, though they are finite. In practice exogenous time limits may be imposed.
(III) Players can in principle "say anything" during their conversational turn, though what they say may very well affect whether their conversation is successful or not.
(IV) While conversations are finite, they may have no designated "last turns;" conversational agents cannot in general foresee who will "have the last word." Hence, people strategize in conversations even when they can't anticipate when the conversation will end, what possible states might arise, or what utterances their opponent will consider.

In order to turn features (I-IV) into a model, we need three things: (i) an appropriate vocabulary of conversational moves for building sequences of message exchanges between players, (ii) goals or winning conditions for conversational players, (iii) a way of modeling the epistemic limitations that players cannot in general foresee the last move of a conversation. Infinitary games like Banach Mazur (BM) games Oxtoby (1957), Grädel (2008), Kechris (1995) furnish a good point of departure, as they reflect some features of (I-IV). For simplicity, we will mostly restrict our attention to two-player win-lose games, allowing us to concentrate on basic conceptual points, though in section 4 we briefly consider extensions. Our theory distinguishes between conversations in virtue of their winning conditions, and different winning conditions require different strategies for achieving them giving rise to different linguistic realizations. We give a precise criterion for the existence of these strategies and a formal model of winning conditions, enabling us to compare different conversational goals and their winning strategies. No extant framework examines the structure of conversations in general and their game theoretic structure in such a precise way.

Our paper is organized as follows. Section 2 introduces the basic points of our model in more detail and considers related work; section 3 gives background on infinitary games and introduces our theory of message exchange games. Section 4 develops a typology of conversations via their winning conditions. We investigate constraints on winning conditions that are intrinsic and normatively necessary for winning conversations most of the time, such as consistency and discourse coherence. These constraint render conversations more complex. Section 5 concludes our paper with some pointers to future work.

## 2 Conversations as infinite games

As stated in the introduction, we think of conversations as games in which the players are trying to achieve a certain end—namely, that the conversation go in a particular way. These games involve a set of sequences of conversational moves and a characterization of winning conditions for players of the game. We now delve deeper into the structure of conversational games. What do they concretely involve? What are the 'moves' of the players, what are their 'strategies' and so on? What are their winning conditions? And how can we model conversational goals in a formal setting?

### 2.1 Signaling games, Grice and opposing preferences

Assuming that conversationalists are rational, what they say and how they interpret what is said should follow as actions that maximize their interests given what they believe. Conversational moves should be calculated via an estimation of best return given what other participants say, which is a natural setting for game theoretic analyses. Game theory has had several applications in pragmatics Lewis (1969), Parikh (1991, 2000, 2001), Benz et al. (2005), Franke (2008), Franke et al. (2009), van Rooy (2003), van Rooij (2004). Much of this literature uses the notion of a signaling game, which is a sequential (dynamic) game in which one player with a knowledge of the actual state sends a signal and the other player who has no knowledge of the state chooses an action, usually an interpretation of the signal. The standard set up supposes that both players have common knowledge of each other's preference profiles as well as their own over a set of commonly known set of possible states, actions and signals. The economics literature contains a detailed examination of signaling games, Spence (1973), Crawford & Sobel (1982), Farrell (1993), Rabin (1990), to name just a few important papers in this area.

Although they have proved useful for many issues, signaling games do not offer a straightforward way to encode the principles we outlined in the introduction, especially for strategic contexts. As a consequence, the model we propose in this paper differs from signaling games in many aspects. However, it does not contradict the predictions of signaling models but rather provides a natural and convenient way of addressing situations that are not transparently expressible as a signaling game's context. We now explain why the strategic contexts considered in this article fall into this category.

A game requires, in order to be a reasonable candidate for modeling non-cooperative contexts, that its structure encodes the players' divergent preferences. As emphasized earlier, the most intuitive way of doing that is to assume a 0 sum game. Signaling games however predict that no communication happens in such games: it can be shown that in equilibrium[2] the sending of any message has no effect on the receiver decision.

An immediate corollary is: assuming that the sender has the possibility of (costlessly) not sending a message and that the sending of any message has at least an infinitesimal cost, $\epsilon$, makes it optimal for the sender to not send anything. This leads to obvious, unintuitive and irrational consequences. Hence, the most straightforward way of setting up non-cooperativity makes communication of any kind is impossible in a signaling game. This means that non-cooperativity of the sort we are interested here should not translate as 0-sum utilities in a signaling model.

Still, there is between perfectly aligned utilities and 0-sum games, a space of games with partially aligned utilities which could encode (some) lack of cooperativity into the context while still allowing for communication to take place. Notice that yielding the right equilibrium is not the only demand to put on the game structure: a precise justification of the chosen utility profile is also needed. In order to use game as part of a general theory of meaning, one has to make clear how to construct the game-context, which includes providing an interpretation of the game's ingredients (types and actions) and explaining why the utility profiles fits the situation to be modeled. Franke (2009), for instance, associates in a principled way an *interpretation game* to a given utterance. Interpretation games form a subclass of signaling models assuming a specific class of sender types actions and preferences. They intend to encode a "canonical context" for an utterance, in which relevant conversational implicatures may be drawn. In interpretation games, the full game structure is determined by the set of sender types: there is a bijection between the set of receiver actions and the set of sender types, and the utility profile is such that both the receiver and sender get rewarded if they coordinate on the sender actual type, and do not gain anything otherwise.

Such a setting is very intuitive and interestingly does not seem to require further precision on what exactly it means for the receiver to take the action $a_t$ associated with receiver type $t$ and why such an action should indeed maximize the receiver payoff if $t$ is the sender's actual type. Even if one can still wonder whether performing $a_t$ means that the receiver believes that the actual state is $t$, or that the receiver is publicly committed to $t$, these interpretations collapse

2 Assuming bounded rationality of conversational agents may restore an effect to messages: for instance the Iterative Best-Response model in Franke (2009) allows a level 2 sender to misdirect a less sophisticated level 1 receiver. However, we are convinced that the conversational examples presented in this article are compatible with a common belief in rationality and require an analysis making such an assumption.

for **Gricean agents**. A Gricean sender should intend to commit to what he believes is true (sincerity), cooperativity should make a Gricean receiver intend to interpret the sender commitment as what the sender intends to communicate, and belief in the sender's sincerity should finally make him believe that the sender believes in what he has committed to. Therefore, the games structure as it stands seems to offer a perfectly adequate level of abstraction.

But things become much more intricate as soon as one is considering potentially non-Gricean players, and this makes the task of understanding and providing justification for a (partially) unaligned utility profile much more involved. It depends on what one takes actions and types to represent. Recall example 2 and imagine, for the sake of argument, that we want to model Bronston's answer with a signaling game involving two sender types: $t_{\mathbf{bank}}$ and $t_{\neg\mathbf{bank}}$, two corresponding interpretative actions $a_{\mathbf{bank}}$ and $a_{\neg\mathbf{bank}}$, and three possible messages, $m_{yes}, m_{no}$ and $m_{company}$. These messages are respectively true in the sets of states $\{t_{\mathbf{bank}}\}$, $\{t_{\neg\mathbf{bank}}\}$ and $\{t_{\mathbf{bank}}, t_{\neg\mathbf{bank}}\}$. Assume also that we want to accommodate a fear of perjury on Bronston's part into the game context. Consider first that performing action $a_{\neg\mathbf{bank}}$ means for the receiver to update his belief to includes that $t_{\neg\mathbf{bank}}$ is the actual sender's type, or at least, to subsequently acts as if it were the case. Under such an interpretation, if a sender sends $m_{no}$ and the receiver takes action $a_{\neg\mathbf{bank}}$, should the sender fear being charged with perjury? Intuitively no, because such an attack would indicate an inconsistent belief of the receiver that $t_{\neg\mathbf{bank}}$ holds (because the action he took is interpreted as such) and does not hold at the same time.[3] Then again, if actions are to be interpreted at the level of public commitments, a receiver who takes action $a_{\neg\mathbf{bank}}$ after receiving $m_{no}$ takes the sender to be publicly committed to $t_{\neg\mathbf{bank}}$, which does not imply that he believes the latter state to be actual, and therefore, a receiver who takes this action is still susceptible to attack Bronston on perjury if he believes that Bronston's actual type is $t_{\mathbf{bank}}$. Bronston's payoff in that case should depend on whether the prosecutor will charge him for perjury and how bad the consequences will be. These considerations illustrate two things: first, in order to deal with non-cooperative contexts, one has to make precise what the exact set of actions is and what they represent—something which may vary according to the nature of the player's objectives (commitments, beliefs, both, something else, ...); second, the payoffs of the sender and receiver may depend on subsequent actions, which requires that the possible outcomes of the

3 We assume here that the prosecutor has an interest to charge Bronston with perjury only if he believes that Bronston actually performed perjury. One can relax this assumption, but that would mean that the prosecutor's beliefs are irrelevant to his subsequent moves and that the commitments-related interpretation of actions should be considered here.

signaling games encode all possible relevant continuations of the conversation. None of this is self-evident and makes a systematic construction of a game context much more difficult than in the cooperative case.

These difficulties stem from the close correspondence between a general formalization of Gricean principles and that of games with shared interests that Asher & Lascarides (2013) establishes. This doesn't entail that in 0 sum games, Gricean principles don't ever apply, but the result does establish that one shouldn't count on Gricean principles as operative; in general one can't assume that players are maximizing quality, quantity or relevance (to one's own conversational ends). Furthermore, there is an important difference between being a non-Gricean speaker, and one that admits to being so. A player's conversational objectives are very likely to include not making such an admission. Conversations can thus exhibit a kind of hide-and-seek game where agents try to expose the "bad" behavior of their opponent while making themselves look good. In example 2, if the prosecutor rejects Bronston's indirect answer, he signals a commitment to Bronston's lack of cooperativity. If Bronston then admit having a swiss bank account, he commits to the fact that he was not cooperative in giving the indirect answer. In such a context, the prosecutor intuitively should claim that Bronston is being non-cooperative but Bronston should try to avoid admitting that he is. In analogous contexts occurring outside of the courtroom, it might be rational for an interrogator who cares for his reputation or his interlocutor's friendship to prefer a misleading answer over formulating a public accusation of non cooperativity that he cannot prove.

Signaling games have difficulties expressing these constraints because of the asymmetry of the sender and receiver in such games. While the sender might reveal something about his type when he sends a message, the receiver obviously does not when he chooses an action. Indeed, in a signaling game with a sender's type $t$ whose interest is to mislead, there is a message $m$ and receiver actions $a$ and $a'$ such that the triple $\langle t, m, a \rangle$ has better utility for the sender than the triple $\langle t, m, a' \rangle$ and that for the receiver the opposite holds, *i.e.* $\langle t, m, a' \rangle$ has a better payoff. Given this we can show that: in any perfect bayesian equilibrium of the game, $m$ is sent with non-zero probability and the receiver uses $a$ has a response to $m$ with a non-zero probability iff there is a sender type $t'$ such that $a$ is a best response to $m$ in $t'$ in the equilibriumf or both the sender and the receiver and the receiver's posterior probability reflects that after the sending of $m$, $t'$ is more likely than $t$. This means that the only basis for a receiver to ever accept a misleading answer is that he judges it more likely that his opponent is cooperative than not cooperative, never that he lacks an argument to confront him, or has reasons to avoid confrontation.

Other models like that of Glazer & Rubinstein (2004) exists that do not use signaling games. However, they also have difficulties in expressing the sort of constraints we have developed above. Signaling games and persuasion games both still take a broadly Gricean view of communication: conversations are essentially information gathering or exchange activities; agents exchange messages for the purpose of affecting the beliefs of the other partner. This is precisely, however, what is in doubt in many conversations. In many conversational settings and in all of our examples, agents converse not in the hope persuading their opponents, but rather to impress or persuade others, and perhaps themselves. Just as Grice captures important aspects of some but not all conversations, people do try sometimes to persuade or to exchange information, but this is not a general framework for all conversations.

We need a different model of conversation. Our players interact with each other and exchange messages that convey objective, public commitments. For instance, D in (3d) commits to Dr. Tzeng's having "torn apart" the nerve by agreeing with LP's description in (3c). D then tries to go back on that commitment in (3g). D may or may not believe this commitment. But if he agrees with (3c), then he is committed to its content, and he can be attacked on the basis of that commitment or subsequent commitments. As our excerpts in our examples make evident, conversationalists often pay careful attention to the commitments of others, not only to explicit commitments but also to their implicatures. For example, in (4) Bentsen seizes on a weak or possible implicature of Quayle's commitments, that he is comparable in Presidential stature to JFK, and attacks Quayle for it. In our examples, the moves players make to defend their commitments or to attack those of an opponent exploit the conventional meanings and even the implicatures that messages have. So our model must enable us to fix the meanings of players' moves to their conventional meaning.

Why do conversationalists make the commitments they do, if they don't do it to persuade their interlocutors or to send a signal that their interlocutors will find credible? Players make the commitments they do, for the purpose of convincing or influencing a third party, which we call *the Jury*. The Jury is for us an abstract role that can be satisfied in diverse ways. In examples (2) and (3), it's the jury of the court; in examples (4) and (6), it was the American electorate. Sometimes the third party may be one of the players, as in example (1) below. The Jury does not as such participate in the conversational game but is rather a scoring function for the game. Players choose their conversational objectives based on what they believe they can defend against their opponents and that will find favor in some way with the Jury. A player attempts to achieve her conversational objective, while her opponent tries to thwart her.

The Jury is an unbiased, rational and competent user of the language of the players and judges on that basis whether a given discourse move or a sequence of moves contributes toward the realization of the conversational objectives of a player or not. We make the simplifying assumption for most of the paper that the Jury can only be convinced by one player.

### 2.2 Why infinite games?

We believe that humans must act as though conversational games were unbounded. If conversations have definite last moves and our players have opposing interests, even the presence of the Jury will not explain why our agents converse in the way they do.

Consider example 2 again, or its non courtroom variant (1). Janet is presented with a Hobbsian choice. Ideally, she would prefer not to answer the question at all or simply lie. To not answer the question or to lie would be rational and what Janet should do, if she were playing a one shot game with no further interaction with Justin (this is akin to the defect move in the Prisoner's Dilemma). As conversations, however, have continuations, many people have the intuition that a refusal to answer will make Janet fare worse in subsequent exchanges. Janet cooperates with her interlocutor in the minimal sense of providing a response to the question, what Asher & Lascarides (2013) call *rhetorical cooperativity*, because of reputation effects. If Janet doesn't cooperate by responding to Justin, she risks receiving uncooperative treatment if in the future she asks a question or make some demands of him. This is a form of the "tit for tat" view of Axelrod (2006). Nevertheless, the reputation argument has its problems. If a conversation is just a finite sequence of one shot games, what holds for a one shot game holds throughout a conversation. Backward induction over such a finite sequence would lead Justin to the conclusion that he should not bother to ask his first question because it is in Janet's to defect at the earliest possible opportunity. If there is a foreseeable last move for one of the players $i$, then she will play to her advantage and defect on the last move, if her opponent has gone along in the discussion. The opponent seeing this will reason by backwards induction to defect at the earliest possible moment. The prediction is that given a foreseeable last move, no message exchange should occur.

If the conversational game is assumed to be infinite, however, the formal argument for the rationality of defection over sequences of exchanges in cases where conversationalists have opposing interests disappears. The argument from backward induction fails because there is no last move from which to begin the induction. However, there is still some explaining to do. A simple “tit for tat” model doesn’t explain why interlocutors cooperate with each other rhetorically, *even if* their roles *vis a vis* their interlocutors are never reversed, even if Janet and Bronston never make any demands of their interlocutors.

In our model, the Jury can force rhetorical cooperativity. A defection will hurt player $i$ if the Jury can infer that $i$’s is defecting because a rhetorically cooperative move would reveal a reason for them not to be persuaded by her. This is also a feature of the model of Glazer & Rubinstein (2001, 2004) but their model is more restrictive. In their model, the Jury only interacts with one sender who must persuade the Jury to accept or reject a message. In addition, the sender of a message is restricted in her choice of messages she can send in a given state, and so the Jury can draw more secure inferences from messages she doesn’t send. Since she can only send certain messages in certain states (e.g., Bronston might be able to say he did not have an account only in a state where he truly does not), a failure to send a message or to respond to a question where the message is directly requested and would be in the player’s interest to send could well indicate that the player is not in the state where such a message is permitted.

We have made no such assumptions about messages, however, because we do not think that messages are tied to states in such a simple way. One can say anything regardless of the state of the world in a conversation. So the reasoning from signals and strategies to the persuasiveness of a player is much more uncertain for the Jury in our games. In addition, while Player $i$ needs to convince the Jury that she has achieved her conversational goals, her goals are more complex than simply getting the Jury to accept the content of a particular message and crucially involve her opponent. Player $i$ could simply refuse to cooperate with her opponent, because she has a general strategy of not revealing information to her opponents. Or she could provide a reasonable defense for why she is not cooperating. In either case, it falls on *the opponent* to make the case to the Jury that player $i$’s lack of rhetorical cooperativity provides a reason to reject $i$’s goals. If attacked, player $i$ can reply to the opponent, defending her lack of cooperativity, and then the opponent must press the issue.

The following excerpt from a press conference by Senator Coleman’s spokesman Sheehan brings out these features of our model. Senator Coleman was running for reelection as a senator from Minnesota in the 2008 US election (thanks again to Chris Potts for this example):

(2) a. Reporter: On a different subject is there a reason that the Senator won't say whether or not someone else bought some suits for him?

b. Sheehan: Rachel, the Senator has reported every gift he has ever received.

c. Reporter: That wasn't my question, Cullen.

d. Sheehan: The Senator has reported every gift he has ever received. We are not going to respond to unnamed sources on a blog.

e. Reporter: So Senator Coleman's friend has not bought these suits for him? Is that correct?

f. Sheehan: The Senator has reported every gift he has ever received. (Sheehan continues to repeat "The Senator has reported every gift he has ever received" seven more times in two minutes to every follow up question by the reporter corps. `http://www.youtube.com/watch?v=VySnpLoaUrI`)

Sheehan, like Bronston, is seeking to avoid committing to an answer to a question. Sheehan's (2b) in response to the reporter's first question could be interpreted as an indirect answer, an answer that implicates a direct answer; the senator didn't comment on the question concerning whether he had received the gift of suits because he felt he had already said everything he had to say about the matter. But in (2) the reporter doesn't accept this rather indirect answer; she says that Sheehan's response was not an answer to her question. In effect, she wants a direct answer to the question concerning the suits. Sheehan then explains why in 2d he won't answer the question. The reporter then presses the issue, and Sheehan becomes rhetorically uncooperative for the rest of the exchange, repeating the same thing. At this point, the Jury will begin to reflect on Sheehan's strategy: is he being rhetorically uncooperative because he has something to hide? His earlier explanation for his defection from rhetorical cooperativity becomes lost, and it becomes more and more plausible that Sheehan won't answer the question because the true answer is damning to his interests. To win given a defection from rhetorical cooperativity, Sheehan has to have a reply for every attack; on the other hand, if the opponent eventually introduces an attack for which the first player does not have a convincing reply (e.g., he simply repeats himself or simply stops talking), the opponent will win.

The need to justify uncooperative moves or defection generalizes. In most strategic situations, in order to win, 0 must engage with questions and remarks of her opponent(s); she must show that her opponent cannot attack her position in such a way that a rational unbiased bystander would find plausible. For any discourse move, we can imagine a potential infinity of attacks, defenses and counterattacks. In successful play, a player has to be able to defend a move $m$ against attacks; she may have to defend her defense of $m$ against attacks and so forth. This is a general necessary victory condition for 0. Let $attack(n, m)$ hold iff move $m$ attacks move $n$; commitments or types of discourse moves that generalize over more specific discourse moves[4] that are used to defend or attack commitments:

**Observation 1** *[NEC] A play is winning for* $0$ *only if for all moves* $n$ *of* $0$ *and for all moves* $m$ *of* $1$*,* $attack(n, m) \rightarrow \exists k(move(0, k) \wedge attack(m, k))$

Conversely for 1, a sufficient condition for winning is the negation of Observation 1. Given (1) 0 wins only if she is prepared for the conversational game never to end and to rebut every attack by 1. It is this constraint that provides a second reason for assuming conversational games to be infinite and is a powerful reason for obeying rhetorical cooperativity.

NEC also has empirical consequences. In virtue of it, we can see why Quayle intuitively loses in example (4). Part of Quayle's winning condition was not to come under attack for his implicit comparison or at least to be able to rebut any attack on his move; that is NEC was also part of his winning condition. But given that he had no rejoinder to Bentsen's unanticipated move, he failed to comply with Observation 1 and so lost.

To Observation 1, we add another, motivated by example (2): to win, 0 should not simply repeat herself in the light of a distinct move by her opponent at least not more than twice.

**Observation 2** *[NR] A play is winning for* $0$ *only if there is no move* $k$ *by* $0$ *such that* $0$ *repeats* $k$ *on* $m$ *successive turns, for* $m \geq 3$*, regardless of what* $1$*'s intervening contributions are.*

NR buttresses NEC's support for rhetorical cooperativity.

While we could weaken the quantifiers in NEC and NR to something like *for most moves of 0*, we are rather interested in the general upshot of such constraints: to model winning play by 0, we need to model a conversation as a potentially unbounded sequence of discourse moves, in which she replies to every possible attack by her opponent. Moreover, at least some of the moves of player

4 Examples of such moves are Answering a question, Explaining why a previous commitment is true, Elaborating on a previous commitment, Correcting a previous commitment, and so on—in fact, they correspond to the discourse relations of a discourse theory Asher & Lascarides (2003).

$i$ must be related to prior moves of her opponent. It follows that it is always risky, and often just rationally unsound, to play a rhetorically uncooperative move like defection without further explanation that is optimal only if it is the last move in a finite game. Defection from rhetorical cooperativity is possible, but it must be explained or defended in any winning play convincing the Jury. A player who plays a rhetorically uncooperative move opens himself up to an attack that will lead to a defeat in the eyes of the Jury, as in (2). That is, relatively weak and uncontroversial assumptions about the beliefs and preferences of the Jury validate rhetorical cooperativity as a component of any winning play.

Even if in practice, conversations don't go on forever, players have to worry about continuations of conversations thus should rationally act as if a conversation were 'potentially infinite'. In such situations, a theory of finite play does not apply and one has to resort to infinite plays. This is why it is necessary to adopt a framework of infinite games. By moving to a framework with unbounded conversational sequences, Aumann & Hart (2003), Aumann & Maschler (1995) show how games with unbounded cheap talk, games involving extended conversations with an infinite talk phase consisting of a pattern of revelations and agreements ending ultimately in an action, make possible equilibria for players that are not available in one shot or even sequences of revelations of bounded length. While we have adopted the simplest of payoff structures for our study, our examples show that unbounded conversational sequences allow players to win conversations that they otherwise couldn't. Had the reporters in 2 been limited to one question and one follow up, they could not have successfully attacked Sheehan in the way they did.

Another reason in favor of using infinitary games is, paradoxically, their simplicity. Given that we cannot impose any intrinsic limitations as to the length of conversations, a formalization of purely finite conversations is more complicated. In an infinitary framework, it is also straightforward to model finite conversations. Finite conversations are not just conversations that stop but crucially involve a point of mutual agreement that the players have finished Sacks (1992). We represent a finite conversation then as one in which a finite sequence terminates with an agreement on a special "stop" symbol that is then repeated forever. More than that, initial prefixes of infinite sequences will play a very important role in the sequel. While our models of conversations will be infinite sequences, all that we ever make judgments on are finite prefixes of such conversations. We will have to evaluate the play of players and whether they have met or are meeting their objectives on such finite prefixes.

## 3 Message Exchange Games defined

We have established that conversations should be modeled as some sort of infinite game. In this section we define such games, which we call *message exchange games*, formally, using Banach Mazur games, a well-known sort of infinitary game, as a departure point and point of comparison. We then make some remarks about the expressive capacities of our new framework and examine how it addresses the problems we found with the signaling game framework.

Let $V$ be a countable, non-empty set. We sometimes refer to $V$ as the vocabulary. For any subset $A$ of $V$, $A^*$ is the set of finite strings over $A$ and $A^\omega$ the set of countably infinite strings over $A$.

**Definition 1** *A Banach-Mazur game (BM game) $BM(V^\omega, Win)$ consists of an infinite set of strings $V^\omega$ together with a winning condition $Win \subseteq V^\omega$.*

The game proceeds as follows. Player 0 first chooses a *non-empty finite* string $x_0 \in V^*$. Player 1 responds by choosing another non-empty finite string $x_1 \in V^*$. Player 0 moves next choosing another finite string $x_2$. This process repeats itself forever yielding a play, an infinite sequence of alternating moves by 0 and 1. Define the flattening $flat$ of a play $p = (x_k)_{k\in\mathbb{N}}$ as the infinite sequence eventually designed by the two players: $flat(p) = x_0 \cdot x_1 \cdot x_2 \ldots \in V^\omega$. Player 0 wins the game if $flat(p) \in Win$. Player 1 wins otherwise. A strategy $f_i$ for player $i$, is a function from the set of finite plays to the set of finite strings, $V^*$. A play $p = (x_k)_{k\in\mathbb{N}}$ of the game, is said to be *consistent* with the strategy $f_i$ iff, for every integer $k$, $k \bmod 2 = i \Rightarrow x_k = f_i(x_0 \cdot x_1 \cdot \ldots \cdot x_{k-1})$. In other words, each move of player $i$ is played according to $f_i$. A strategy $f_i$ is said to be winning iff in every play consistent with $f_i$, player $i$ wins.

BM games suggest a natural model for conversations: participants alternate turns in which they utter finite contributions. These contributions add to each other, and together form a conversation. This process potentially goes on indefinitely, or, at least strategic reasoning requires thinking of it that way. However, BM games "erase" the information of who said what in the following sense:

**Proposition 1** *Let $BM(V^\omega, Win)$ be a BM-Game. Then, for any play $p$ and every play $p' \in flat^{-1}(flat(p))$, player $i$ wins in $p$ iff player $i$ wins in $p'$.*

Given any infinite sequence $s$, any infinitely countable set $turns \subseteq \mathbb{N}$ such that $min(turns) > 0$ yields a play in $flat^{-1}(s)$ and conversely: every element in $turns$ specifies a position in $s$ which is the end of a player's move. A corollary of the above proposition is that we cannot define a winning condition that imposes for instance that player 1 says something in particular, as long as she and 0 don't infinitely repeat the same single move. We formalize this observation as follows:

**Corollary 1** *Let $BM(V^\omega, Win)$ be a BM-Game. Let $y \in V^*$ be a finite sequence such that there is at least one infinite sequence $w$ in $Win$ such that $w \notin V^* y^\omega$ or $|y| > 2$. There is a play $p$ of the game such that $y$ is never a substring of any move of player 1 in $p$.*

The idea for the proof of this corollary is simple: since $w$ does not end with infinite repetitions of $y$, every occurrence of $y$ in $w$ is eventually followed by something which is not $y$, call it $x$. It suffices to define the alternation of turns so that $x$ constitute exactly the turns of 1. More formally: let $i$ be a position in $w$ at which $y$ appears. define $l_i$ as $Min(\{l \in \mathbb{N} \mid y$ does not occur at $i+l \times |y|$ in $w\}$. $l_i$ exists by hypothesis since otherwise $w$ would end with infinite repetitions of $y$. For any position $k$ in $s$, let $n_y(k)$ be the position of the first occurence of $y$ in $w$ after $k$. Define inductively $turns$ with $n_y(0) + l_{n_y(0)} \times |y| \in turns$, and for any $k$ in turns, $n_y(k) + l_{n_y(k)} \times |y| \in turns$. $turns$ yields a winning play in $flat^{-1}(w)$ for which player 1 never "says" $y$.

BM games have a limitation that require us to introduce a more structured type of game given our principle I. A given conversationalist might have as a goal that her interlocutor *and only her* commit to a particular content, or answer a particular question, which BM games do not allow. Consider again (2). There's an important difference between Bronston's response to a question by the prosecutor and the prosecutor's offering that information himself, and that difference can't be captured in BM games under all interesting scenarios. Discourse moves contain more information than the sentence itself. The discourse move that Bronston commits to a negative answer to (2a) provides more information than just the string *no sir* provides, and it is such moves that are of interest.

To remedy the expressive limitations of BM games, we introduce our variant, message exchange (ME) games. An ME game involves a vocabulary of discourse moves $V$ (which for the moment we consider as just atoms) that yields two disjoint sets $V_0 =_{def} V \times \{0\}$ and $V_1 =_{def} V \times \{1\}$, standing for the respective vocabularies of 0 and 1. The game is played with 0 and 1 alternatively choosing finite sequences **in their own vocabulary** $V_0$ or $V_1$. We can take the vocabulary of the players to be common as in the above definition or to differ; the game is structured as to encode the information on which player played what into the plays so that we can determine whether 0 or 1 commits to a particular move. A turn by $i$ is an element of the set $(V_i)^+$.

The plays of the ME game over $V$ are $\omega$ sequences in $(V_0 \cup V_1)^\omega$.

**Definition 2** *An ME game is a pair $G = ((V_0 \cup V_1)^\omega, Win)$, with $Win \subseteq (V_0 \cup V_1)^\omega$.*

Player $i$, at turn $j$, picks a *n*on-empty finite sequence of moves forming a finite path in $V_j$, which we interpret as a finite (dynamic) conjunction of move formulas. At any point in the conversation, these finite sequences of moves concatenate and give us a finite conversational play or path in $(V_0 \cup V_1)^*$.

An important condition on $Win$ is whether it hinges on which of the players made a particular discourse move. We call such winning conditions *decomposition sensitive*. Let $\pi$ denote the natural projection of $V_0 \cup V_1$ onto $V$ ($\pi(v, i) = v$). Define $\pi_\omega$ as the extension of $\pi$ into a projection of sequences in $(V_0 \cup V_1)^\omega$ onto $V^\omega$: $\pi_\omega((v_k)_{k\in\mathbb{N}}) = (\pi(v_k))_{k\in\mathbb{N}}$ (where all the $v_k$ belong to $V_0 \cup V_1$).

**Definition 3 (Decomposition sensitive winning conditions)** *$Win \subseteq (V_0 \cup V_1)^\omega$ is* decomposition sensitive *iff $\exists W \subseteq \pi_\omega(Win)$ such that $\neg(\pi_\omega^{-1}(W) \subseteq Win)$.*

Conversely, an ME game $G$ with a *d*ecomposition invariant $Win_G$ is one where: $\exists W \subseteq V^\omega$ such that a sequence $\sigma_1 \in (V_0 \cup V_1)^\omega$ is an element of $Win_G$ iff $\pi_\omega(\sigma_1) \in W$. Thus, if 0 has a winning strategy in $G$, she also has one for attaining $W$ in the BM game $BM(V, W)$, and conversely, if she has a winning strategy for attaining $W$ in $BM(V, W)$, there is a sequence of plays that she can make regardless of what 1 does that will guarantee her a sequence $s \in W$. That sequence of plays yields a sequence $\sigma \in (V_0 \cup V_1)^\omega$ in which 0 and 1 are assigned different contributions at turns such that $\pi_\omega(\sigma) = s$ and so $\sigma \in Win_G$. Thus, if 0 has a winning strategy in $BM(V, W)$, she also has a winning strategy in $G$. We have thus shown:

**Proposition 2** *Given an ME game $G = ((V_0 \cup V_1)^\omega, Win_{ME})$ where $Win_{ME}$ is decomposition invariant, 0 will have a winning strategy in $G$ iff she has a winning strategy in the BM game $BM(V^\omega, Win_{BM})$ over $V^\omega$ where $Win_{BM} = \pi_\omega(Win_{ME})$.*

In other words, when ME games involve decomposition invariant winning conditions, they collapse to BM games, and the existence of a winning strategy is predicted by the basic theorem for BM games, which we discuss in the next section.

### 3.1 The Jury, constraints and meanings

But first we revisit a problem we posed for signaling games that define content in reflective equilibrium, the problem that signals have no meaning in conversations where players' interests are opposed. In our model of conversation, the opposing interests of the players do not impede communication of content but rather presuppose a set content; both players have to have a clear and defensible idea of what their opponent has committed to if they hope to win a message exchange game. While our ME games allow a player to say anything on her turn, just saying anything lacks certain important elements intrinsic to a good or winning play, and these elements end up determining this content. Players set their winning conditions vis a vis an audience that makes requirements on winning conversations. And so we need to give a more detailed model of the Jury.

The Jury can either be biased towards a particular victory condition that player 0 must guess or unbiased and accept whatever victory condition 0 chooses to play. In either case it is swayed by argument and verifies whether a particular victory condition has been met. The Jury rates each contribution by a player in individual turns or small sequences of turns with respect to whether they get a player closer to a given goal or make it more difficult to attain. We will suppose that turns are evaluated as either helping 0 achieve a particular goal, hindering her or having no effect via a function $\|t\| \in [-1, 1]$. For the moment we will assume that the Jury is unbiased and so the Jury's and 0's conception of the winning condition coincide.

The function $\|.\|$ should also verify necessary conditions on good discourse like consistency and coherence, which we now describe. Discourse consistency can be defined in different ways, but it must respect the rules of valid inference. In our ME games, the rules of inference for the logical connectives and quantifiers, as well as the conventional lexicon for non logical terms, impose a notion of consistency on play. If 0, for instance, maintains in example 1 that she proved a theorem but also that she did not prove it, she is inconsistent and the Jury will conclude she is confused. If she claims that she proved the theorem but also that if the theorem has a proof, it hasn't been found yet, she is also inconsistent. Such inconsistency precludes her from her winning condition. We can further extend our notion of consistency by supposing further that our games *are situated* in the sense that they involve deictic reference to non-linguistic objects and properties like natural kinds. Courtroom cases typically involve extra-linguistic elements fixing the meaning of certain referring terms (imagine introducing pictures in a courtroom of a particular character, or the character himself). So consistency will involve more complex rules like how to adjust one's commitments in the light of new evidence about a natural kind or about an individual. Any violation of consistency will lead to an immediate attack by the opponent —*you just contradicted yourself so how can we (i.e. the jury) believe anything you're saying.* This is the sense of LP's closing comment in example 3. So requiring that a player $i$'s winning plays form a consistent set of formulas in $V_i$ will place constraints on the meanings of her expressions.

For the time being, we consider the following very simple form of consistency, though ideally a more sophisticated form is needed.

**Definition 4** *A play $p$ of an ME game over vocabularies $V_0$ and $V_1$ is consistent for player $i$ iff $p \upharpoonright V_i$ does not both contain $\phi$ and $\neg\phi$ for any formula $\phi \in V_i$.*

How does consistency affect the Jury? If an agent $i$ makes inconsistent contributions, then her contributions automatically entail that the victory conditions of the other agent have been achieved. Given that $i$ is inconsistent, her contributions entail a commitment to any content whatsoever and no information anymore; she commits to any finite sequence of discourse moves on every turn, and so $j$ just needs to make the appropriate moves on each turn to achieve her winning condition.

Successful play also involves rhetorical cooperativity. Defection from the conversation is not usually a winning option for either player. The space of possibly coherent attacks on a message places constraints on the meaning of messages. An attack, or in discourse terms a Correction Asher & Lascarides (2003), can apply in principle to practically any word in a player's contribution, as the following adaptation of a famous example of Strawson's shows:

(3)
- a. 0: A man fell off a bridge.
- b. 1: It wasn't a man. It was a woman, and she didn't fall; she was pushed.
- c. # 1: No, John is a basket ball player.

However, not just any move can count as a coherent attack. 3c is an incoherent discourse move and cannot be interpreted as a sensible attack or any other discourse move relating to the claim in 3a. Though player 1 can make the move in 3c, it won't do him any good vis a vis the jury, and an opponent can attack it successfully as an incoherent move. Competent speakers of a language can tell quite well when an attack is coherent or not and so our players will also have to play only coherent attacks if they wish to succeed in their winning condition. Our model of the Jury below reflects this in its estimation of the type of each player; a player who is successfully attacked will suffer in the Jury's estimation of her type.

Attacks aren't the only sort of discourse move that we have to countenance. There are also rebuttals, defenses or supporting moves for claims and many others. All of these moves have coherence requirements. This is what discourse theories study, they make use of the lexical and compositional content of their relata to infer such relational discourse moves and check their coherence Asher & Lascarides (2003). Requiring winning plays to involve coherent discourse moves constrains the meaning a message can have.

In effect our language $V$ has a rich structure, borrowed from discourse theories like Asher & Lascarides (2003). We have descriptions of contents of *discourse constituents* (which one can think of as elementary discourse moves) and these discourse constituents are arguments to various discourse relations like Question Answer Pair and Correction. We thus have a countable set of discourse constituent labels $DU = \{\pi, \pi_1, \pi_2, \ldots\}$, and a finite set of discourse relation symbols $\mathcal{R} = \{R, R_1, \ldots R_n\}$. Our vocabulary $V$ consists of formulas of the form $\pi\colon \phi$, where $\phi$ is a description of the content of the discourse unit labelled by $\pi$ and $R(\pi, \pi_1)$, which says that $\pi_1$ stands in coherence relation $R$ to $\pi$. Following Asher & Lascarides (2003), each discourse relation symbolized in $V$ comes with constraints as to when it can be coherently used in context and when it cannot. It is these constraints that give the meanings of agents' messages and of their commitments, *irrespective of what they believe about the contents of those messages.*

With our vocabulary $V$ now fixed, we can specify rhetorical cooperativity more precisely by noting that a sequence of conversational moves can be represented as a graph $(DU, E_1, \ell)$, where $DU$ is a set of vertices each representing a discourse unit, $E_1 \subseteq DU \times DU$ a set of edges representing links between discourse units that are labeled by $\ell : E \to \mathcal{R}$ with discourse relations. We can now define rhetorical cooperativity using the following two concepts. Let $\mathcal{T}$ be the set of turns in an ME game and let $\rho_i : \mathcal{T} \to \wp(V_i)$ be the projection from a turn to the set of DU's of the contribution by $i$ therein.

**Definition 5 (Coherence)** *A contribution by $i \in \{0, 1\}$ is coherent on turn $\tau$ if for all $\pi \in \rho_i(\tau)$ there exists $\pi' \in (\rho_i(\tau') \cup \rho_{1-i}(\tau'))$ where $\tau'$ is $\tau$ or some previous turn such that there exists $e \in E_1$ such that we have $(e(\pi', \pi) \vee e(\pi, \pi'))$ holds.*

**Definition 6 (Responsiveness)** *Player $i \in \{0, 1\}$ is responsive on turn $\tau$ if there exists $\pi \in \rho_i(\tau)$ such that there exists $\pi' \in (\rho_{1-i}(\tau'))$ where $\tau'$ is the previous turn such that for some $e \in E_1$ we have $e(\pi', \pi)$.*

**Definition 7 (Rhetorical cooperativity)** *Player $i$ is rhetorically cooperative in an ME game $G$ if every turn in $G$ by $i$ is both coherent and responsive. $G$*

*is rhetorically cooperative if the play of both players is rhetorically cooperative.*

Note that one can coherently continue at turn $k+2$ one's contribution from turn $k$, ignoring the opponent contribution in turn $k+1$. Responsiveness on the other hand, forces a player to address the opponent's last turn in some way (acknowledging, correcting, answering questions, ...).

Recall our constraint NEC, a condition according to which to win a player must be able to respond to every attack. We can characterize strings that provide attacks and responses effectively in our discourse language Venant et al. (2014), and we can characterize NEC as follows:

**Definition 8 (NEC)** *NEC holds for Player $i \in \{0,1\}$ on turn $\tau$ iff for all $\pi' \in (\rho_i(\tau') \cup \rho_{1-i}(\tau'))$ where $\tau'$ is some previous turn, there exists $\pi \in \rho_i(\tau)$ and there is an $e \in E_1$ such that $e(\pi', \pi)$, Attack$(\pi')$ and Response$(\pi)$.*

Now let's see how the Jury enforces these constraints. The Jury will penalize contributions by that are not coherent, and it will penalize a player that is not responsive on her turn. While being incoherent or unresponsive on a turn is not a game changer; being inconsistent is—inconsistency makes the player automatically lose. In addition, our Jury is sensitive to attacks that it deems successful; and it is sensitive to ones with no reply—i.e. ones that do not conform to NEC. The following model of the Jury with two components makes these claims concrete:

The Jury assigns a rating to the contribution in turn $\tau_k$ with the following constraints:

- if the player 0 in $\tau_k$ fails to respect coherence in $\tau_k$ then $coh_0(\tau_K) = \begin{cases} -1 \text{ if } k \,\texttt{mod}\, 2 = 0 \\ 1 \text{ otherwise.} \end{cases}$
- if the player of $\tau_k$ is not responsive in $\tau_k$, then $res_0(\tau_K) = \begin{cases} -1 \text{ if } k \,\texttt{mod}\, 2 = 0 \\ 1 \text{ otherwise.} \end{cases}$.
- If 0 is inconsistent by turn $\tau_k$ of $\sigma$ and 1 is not, then $cons_0(\tau_{k'}) = -1$ for all $k' \geq k$. Otherwise, $cons_0(\tau_{k'}) = 0$
- In all other cases where a winning condition *Win* distinct from coherence, consistency, responsiveness and nec, $win_0(\tau_k) = 1$ if $\tau_k$ advances 0 toward *Win*; $Win_0(\tau_k) = -1$ if $\tau_k$ takes 0 further away from *Win*; $Win_0(\tau_k) = 0$ otherwise.

The Jury also maintains a probability distribution over to types: $\text{BAD}_0$ and $\text{GOOD}_0$ modeling the gain or loss of credibility that 0 has faced so far. At each turn we write this probability as $P_k$, and it is defined as follows:

- $P_0(\text{GOOD}_0) = 1$ and $P_{k+1}(\text{GOOD}_0) = P_k(\text{GOOD}_0|\sigma_k)$, where $\sigma_k$ is the initial sequence of $k$ turns in the game.
- $P_k(\text{BAD}_0) = 1 - P_k(\text{GOOD}_0)$.
- if 1 successfully attacks 0 at turn $k$, then $P_k(\text{GOOD}_0) = P_{k-1}(\text{GOOD}_0|\sigma_k) = c_k P_{k-1}(\text{GOOD}_0)$ where $0 \leq c_k < 1$ is a constant representing the severity of punishment per single move of a player $i$ by the jury ($c_k = 2/3$ is a good example).
- Conversely, if 0 successfully attacks 1 at turn $k$ (this includes a good response to an attack move by 1), then $P(\text{BAD}_0|\sigma_k) = c_k.P_{k-1}(\text{BAD}_0)$.

These two ingredients contributes to a definition of the Jury's evaluation in the following way:
$\|\tau_k\|$ of the $k^{th}$ turn's benefits to 0:

$$\|\tau_k\| = coh_0(\tau_k) + res_0(\tau_k) + con_0(\tau_k) + P_k(\text{GOOD}_0)(win_0(\tau_k))$$

$$\|\sigma\|_0^{\uparrow} = \liminf_{n \to \infty} \left[ \frac{1}{n} \sum_{k=1, \tau_k \in \sigma}^{n} \|\tau_k\| \right]$$

Then 0 obeys the *Jury condition* for a sequence $\sigma$ only if $\|\sigma\|_0^{\uparrow} > 0$ and 1 wins otherwise. It is easy to see the following:

**Proposition 3** *If a necessary condition on Win is the Jury condition, then to win* 0 *must respect consistency, and must satisfy NEC, responsiveness and coherence more often than not..*

The fact that the Jury enforces our constraints implies that the meaning of a move is largely fixed by its consistent and coherent uses in context, how it can be attacked and how it can be defended, amplified on and so on. This is a counterpart of a well-known fact about formal languages: the models of a formal language are determined by the consistency notion of the language's underlying logic or semantics. Because our players play infinite sequences, a player can completely specify a model for a countable first order language using Lindenbaum's procedure.[5] Suppose player $i$ plays a move $\phi$. Using the consistency and coherence requirements on winning plays, she can build a

5 See e.g. Chang & Keisler (1973).

maximal consistent and coherent set of formulas from $V$ by adding a consistent formula with a coherent relation at each turn to what she has already played. Since $V$ contains at most $\omega$ many formulas, an infinite play of an ME game suffices to construct a maximal consistent and coherent set.[6] In fact the space $(V_0 \cup V_1)^\omega$ contains *all* maximal consistent sets for the language. Thus:

**Proposition 4** *Let $V$ be a countable first order language. Then there is a set $C \subseteq (V_0 \cup V_1)^\omega$ that consists of all plays that are consistent and rhetorically coherent and that specifies all the intended models of $V$.*

Not every play specifies a full model. However, the game structure itself does. And in our idealized setting, the plays that are consistent and specify models are common knowledge of the participants and of the Jury.

Our ME games thus enforce an exogenously specified notion of meaning, specified by linguistic theory. This includes implicatures, which would seem to mark an important difference with the signaling models of section 2.1). In signaling models, implicatures arise as a byproduct of cooperativity in the game's equilibrium; our model takes them to be provided by linguistic theory, and then predicts agents' attitude toward them in the conversation's continuations. We could thus think of a signaling model as one of implicature generation, while our model is one of implicature "survival".

We think the situation is more complicated for two reasons. First, on a commitment-based view such as ours, the constraints of consistency and coherence determine implicatures. For instance, Quayle's implicature $K$ in (4) that he was comparable to John Kennedy as a politician translates in our model into the fact that it is consistent and coherent both for Bentsen to commit that Quayle committed to $K$ and for Quayle to commit that he did not commit to $K$. Now as noticed earlier, a decision to exploit or to deny an implicature brings with it a commitment that the linguistic premises (cooperativity, sincerity, competence...) of the implicature's derivation hold, or do not hold. This being understood, it does not matter whether one implements those premises within a logical theory, or within a signaling game's utility profile. But linguistic constraints like coherence and consistency do further work. Consider, example (2) with a fully cooperative Bronston. Where does the implicature to the "No" answer comes from in the first place? This implicatures is fundamentally tied to coherence. Inferring "Yes" through Bronston indirect answer makes him less coherent than inferring "No", because the "No" allows for an implicit contrastive discourse relation ("No I did not. [But] the company had one" while the yes would require an explicit marker "the company had one **too**" to

6 Or a maximal consistent and saturated set.

infer a relation (this is moreover actually confirmed by the natural prosody of Bronston's answer). Any model would have to rely one way or another on a pragmatic theory to explain utilies and model this assymetry.

Second, we do not think that signaling games are independent of an exogenous linguistic theory with regard to implicatures. One of the main concerns of the model in Franke (2009) is to bring conventional meaning back into signaling model, which is not innocuous by constraining the set of player types in the game. In order to capture implicatures properly, one needs to make conventional meaning a part of the signaling model. To this end Franke (2009) suggests that the set of types constitute a potential answer set to the question under discussion; hence, determining the game's context requires a pragmatic model as well.

In conclusion, both signaling games and ME game need to appeal to an exogenous theory. Signaling games give a nice implementation of the Gricean theory where linguistic considerations can often be "hidden" into the game context, whereas ME games allow a higher level form of quantification over those possible game contexts, which is crucial to account for the possibility of Gricean or non-Gricean speaker.

### 3.2 Some interim remarks on ME games

BM games are classically played by 2 players, as are ME games. However, conversations are not limited to two players only but may involve several players. Most of our examples in Section 4 are of that nature. Our model can accommodate such a scenario as follows: when a conversation involves $n$ players ($n > 2$ say), the player whose objective is to achieve the winning condition *Win* is taken as Player 0 and all the other players 'together' form Player 1. Any and everything of what is said by these $n - 1$ players constitutes Player 1 moves in the conversational ME game. Such an assumption is standard in the theory of multiplayer games.

In ME games, winning conditions are defined over sets of infinite sequences. However, our Jury witnesses actual conversations that are perforce only finite, initial prefixes of such sequences and forms a judgment concerning winners and losers of conversation on this basis. So our theory must provide a means for verifying whether a winning condition that applies to infinite strings holds or not in virtue of finite prefixes of those strings when possible. ME games whose winning conditions are not finitely verifiable will give rise to actual conversations for which a winner cannot once and for all be determined.

It may seem odd that we define a conversation's winning condition for a player using only sequences of discourse moves involving commitments. Don't agents engage in conversation typically to get their interlocutors or observers of the conversation to do some non-linguistic action? As our vocabulary can be what we can like, we can add to our vocabulary of discourse moves descriptions of non-linguistic actions or states, like *player 1 buys the goods*, which is a move that player 1 might make at the close of a negotiation. In principle we can include a description of whatever actions that are pertinent to winning conditions in a conversation.

ME games involve separate vocabularies for our two players. We have assumed that the same types of move from a vocabulary $V$ are available to both, but our games distinguish which player plays which move by restricting 0 to play from $V_0 = \{(v, 0)|v \in V\}$ and 1 to play from $V_1 = \{(v, 1)|v \in V\}$. However, players may play with *different* vocabularies, for instance where 0 plays from the set $\{(u, 0)|u \in U\}$, 1 plays from $\{(v, 1)|v \in V\}$ and $U \subsetneq V$. That is, the moves envisioned by one player is a strict subset of the other player's set of envisioned moves. We plan to investigate this possibility in a future paper.

It's also possible that as in chess, one player, indeed both players, might be playing more than one ME game, with the same vocabulary of moves. The possibility of using multiple ME games seems an intuitive way to analyze the complex goals of strategic conversationalists like debaters, who have both positive goals and also goals to thwart the positive goals of the opponent. With two ME games, a debater could play 0 in one game and 1 in the other. Alternatively, given the possibility of a biased Jury, 0 may be playing one ME game according to her conception of the Jury, while the Jury takes her to be playing another. Indeed, the interactions between players and the Jury offer a whole range of options, that plan to explore in the future.

## 4 Winning conditions

Let's recap. We've argued that the analysis of strategic conversation requires a different framework with novel features, from those used in most game theoretic analyses of conversation. We've established five necessary constraints on winning conditions in virtue of a model of the Jury—consistency, coherence, NEC from observation 1, NR from observation 2 and responsiveness; and we've shown that these constrain the meaning of the signals players use independently of the players' beliefs or preferences, in contrast to other game theoretic frameworks.

Nevertheless, while our model of the Jury evaluates individual contributions of players it does not completely determine winning conditions or the 'shapes' of conversations that depend on them. This is something we now investigate.

We will characterize precisely the conversational objectives that players choose, how complex they are, whether there is a winning strategy in the game for achieving them, and, if there is a winning strategy, how complex it is. In order to do this, we need a natural way of classifying sets of infinite strings that satisfy different conversational goals. The fact that the sets of infinite strings in BM games and ME games define a topology will allow us to make use of the Borel Hierarchy, a natural measure of topological complexity, to characterize different types of winning conditions. The place of a winning condition in the Borel Hierarchy in turn determines the complexity of the winning strategy that a player should employ.[7] We will examine the structure of the winning set *Win* in a ME game when *Win* lies in the low levels of the Borel hierarchy.

To define the topology on the infinite strings in our ME games, we first need to fix our domain $D$, which is the set of all possible plays in the ME game $G = ((V_0 \cup V_1)^\omega, \mathit{Win})$. Note that $D = (V_0 \cup V_1)^\omega \setminus (V_0^\omega \cup V_1^\omega)$. That is, since the players have to play *finite and non-empty* sequences, this excludes $(V_0^\omega \cup V_1^\omega)$ from the set of possible plays. We define the open sets in our topology to be sets of the form $A(V_0 \cup V_1)^\omega \cap D$ where $A \subset (V_0 \cup V_1)^*$. We shall often denote this set as $\mathcal{O}(A)$. Intuitively, $\mathcal{O}(A)$ denotes all possible ways in which a play can continue after a string $u \in A$ has been played. If $A$ is the single string $\{u\}$ then we shall abuse notation and write $\mathcal{O}(u)$ instead of $\mathcal{O}(\{u\})$. The closed sets are as usual the complements of the open sets. Suppose for example, (2) provides an initial segment of an ME game. $\mathcal{O}(2)$ is all the ways that the conversation can continue. This is an open set over the set of plays $D$ in the ME game.

Using this topology, we inductively define the Borel sets, $\Sigma^0_\alpha$ and $\Pi^0_\alpha$ for $1 \leq \alpha < \omega_1$. Let $\Sigma^0_1$ be the set of all open sets. $\Pi^0_1 = \overline{\Sigma^0_1}$, the complement of the set of $\Sigma^0_1$ sets, is the set of all closed sets. Then for any $\alpha > 1$ where $\alpha$ is a successor ordinal, define $\Sigma^0_\alpha$ to be the countable union of all $\Pi^0_{\alpha-1}$ sets and define $\Pi^0_\alpha$ to be the complement of $\Sigma^0_\alpha$. $\Delta^0_\alpha = \Sigma^0_\alpha \cap \Pi^0_\alpha$. Below is a schematic picture of the initial sets and their inclusion relations in the Borel Hierarchy.

$$
\begin{array}{ccccccccccc}
 & & \Sigma^0_1 & & & & \Sigma^0_2 & & & & \Sigma^0_3 \; \text{- - -} \\
 & / & & \backslash & & / & & \backslash & & / & \\
\Delta^0_1 & & & & \Delta^0_2 & & & & \Delta^0_3 & & \\
 & \backslash & & / & & \backslash & & / & & \backslash & \\
 & & \Pi^0_1 & & & & \Pi^0_2 & & & & \Pi^0_3 \; \text{- - -}
\end{array}
$$

Similar to the above, let us define a topology on $V^\omega$ by specifying the open sets as $AV^\omega$ where $A \subset V^*$. We note that since our "flattening" or projection function $\pi_\omega \colon ((V_0 \cup V_1)^\omega \to V^\omega$ is an onto homomorphism, the Borel

7 See, for instance, Revalski (2003-2004) for a nice survey on infinitary games.

complexity of $\pi_\omega^{-1}(A)$ = the Borel complexity of $A$, where $\pi_\omega^{-1}(A) = \{s \in (V_0 \cup V_1)^\omega | \pi_\omega(s) \in A\}$. Further, $\pi_\omega$ maps open sets in $(V_0 \cup V_1)^\omega$ to open sets in $V^\omega$. However, closed sets in $(V_0 \cup V_1)^\omega$ maybe mapped either to closed sets or to open sets in $V^\omega$. By a simple inductive argument we can show that the Borel complexity of sets do not increase under the map $\pi_\omega$.

For ME games with decomposition invariant winning conditions, which are just BM games, the Banach-Mazur theorem states necessary and sufficient conditions for the existence of a strategy to achieve *Win*. The theorem intuitively says that Player 0, the player who starts the conversation, can win if her strategy takes into account, or has an 'answer', for almost all possible situations when her turn to speak may come. That is, the set of situations that her strategy doesn't take into account must be 'small' in a sense that we define below. Example 5 from the introduction illustrates the problem. Feynmann, who confronts his students eager to stump him by beginning the conversation with "So what shall we discuss today?", automatically throws the floor open for all possible topics that might arise. He can achieve his objective only if the continuations of this initial question that do not fall within his winning condition are very few. He must have a convincing response to any possible question or topic that may be thrown at him that enables him to get to his chosen topic.

To understand this theorem, we need some definitions. A set is nowhere dense if its closure contains no non-empty open set. A set is meager if it is a countable union of nowhere dense sets. Meager sets represent sets which are 'small' in a topological sense. The complement of a meager set is a co-meager (or topologically 'large') set. A topological space is called a Baire space if the countable intersection of dense sets is dense. That is, every meager set is nowhere dense. We will work in a Baire space.

**Theorem 1 (Banach-Mazur (Mauldin 1981))** *Given a BM game $BM(V^\omega, Win)$, (i) Player 1 has a winning strategy if and only if Win is a meager set; (ii) Player 0 has a winning strategy if and only if, there exists a finite string $x$ such that $O(x) - Win$ is meager (that is, Win is co-meager in some basic open set).*

### 4.1 Winning conditions in conversations: Reachability and Safety

Let's start by looking at the winning condition for the conversation in Example 1, which, at least on a certain interpretation, is a very simple, decomposition invariant condition. Suppose that in order to achieve her winning condition, candidate A has only to mention at some point, it doesn't matter when, the theorem that she has proved (leaving aside the constraints NEC, NR, consistency or discourse coherence, which make winning conditions more complex). In other words, for A to win the game, her conversation $x$ must eventually contain this move. More generally, conversations in which the objective of a player is to simply "touch upon" a certain topic exhibit the following shape:

$$Win = Reach(R)$$

Reachability, a characteristic property of many conditions classified as $\Sigma_1^0$ in the Borel hierarchy, is defined as follows. Given a non-empty subset $R \subset (V_0 \cup V_1)$ of the elements of the vocabulary, a string $x$ in $(V_0 \cup V_1)^\omega$ is said to reach $R$ if the elements from $R$ occur somewhere in $x$. More formally, for a string $x$ over the vocabulary $(V_0 \cup V_1)$ we let $x(i)$ denote the $i$th element of $x$. We define $occ(x) = \{a \in (V_0 \cup V_1) \mid \exists i, x(i) = a\}$ to be the set of all the elements of $(V_0 \cup V_1)$ which occur in $x$. Then

$$Reach(R) = \{x \in D \mid R \subseteq occ(x)\}$$

is the set of all strings in which the elements of $R$ occur at least once. The reachability set $R$ of example 1 is just a singleton. Since this winning condition is decomposition invariant, by the BM theorem A has a winning strategy, since the winning condition picks out an open set of strings of discourse moves.

Just as Reachability characterizes many conditions in $\Sigma_1^0$, Safety characterizes an important subset of these $\Pi_1^0$ conditions and is defined as follows. Suppose $S$ (the 'safe' set) is a subset of $(V_0 \cup V_1)$.

$$Safe(S) = \{x \in D \mid occ(x) \subset S\}$$

is the set of all strings which contains elements from $S$ alone. That is, the strings remain in the safe set and do not move out of it. One common sort of $\Pi_1^0$ condition is to prevent player 0 from reaching a $\Sigma_1^0$ condition; another is to avoid a certain commitment or finite set of commitments.

**Note** An alternative way of thinking about reachability and many $\Sigma_1^0$ conditions is to look at their definitions in terms of temporal logic. As our space of infinite strings is a set of linear orders, formulas of linear temporal logic (LTL) can describe some of its subsets, in particular many reachability conditions. For any element $a \in (V_0 \cup V_1)$, let the proposition $p_a$ denote the property of visiting or playing $a$. For some finite $R \subset (V_0 \cup V_1)$, the LTL defining formula for the strings that reach $R$ is:

$$\phi_{reachable(R)} = \bigwedge_{a \in R} \Diamond p_a$$

where $\Diamond$ is interpreted as *eventually*.[8] A reachability formula of the form $\Diamond p_a$ is true at an index $i$ of a sequence $x$ ($x, i \models \Diamond p_a$) iff for some $j \geq i$, $x, j \models p_a$. A string $x$ satisfies $\Diamond p_a$ ($x \models \Diamond p_a$) iff at the initial point 0 of $x$, $x, 0 \models p_a$. Safety also has an LTL defining formula for the strings that stay in $S$: where $\Box$ is interpreted as *always*,

$$\phi_{safe(S)} = \Box \bigvee_{a \in S} p_a$$

For a safety goal of the form $\Box\phi$, $x, i \models \Box\phi$ iff for all $j \geq i$, $x, j \models \phi$. $x \models \Box\phi$ iff $x, 0 \models \Box\phi$.

The simple $\Sigma^0_1$ winning condition of candidate A's is decomposition invariant. However, this is not so for other goals. Consider (1) again. Justin is playing a game with a disjunction of reachability conditions as his winning condition: his goal is to get Janet either (i) to admit that she has been seeing Valentino or admit that she hasn't. These characterize an open set in an ME game where 0 is Justin. Nevertheless, unlike candidate A's, Justin's winning condition depends on Janet's making a certain commitment, a decomposition sensitive winning condition. We need to analyze the topological characteristics of such winning conditions.

We first look at winning conditions that are decomposition sensitive in a particular way: *Win* depends on 0's making a particular contribution at each turn. We call these conditions *rhetorically decomposition sensitive*. Our constraints of responsiveness, coherence, consistency and NEC are all rhetorically decomposition sensitive conditions.

**Proposition 5** *If Win is rhetorically decomposition sensitive, then it is meager.*

**Proof** Let $x \in Win$ be a winning play. Since *Win* is decomposition sensitive, for every prefix $x_n$ of $x$ which ends with a contribution of 1 there exists a finite $y_n$ such that $\mathcal{O}(x_n y_n) \cap Win = \emptyset$. Since $x$ was arbitrary, this means the closure of *Win* cannot contain a dense open set. Hence, *Win* must be meager. $\Box$

Given the constraints enforced by the Jury, we will be mostly interested in rhetorically decomposition sensitive winning conditions in ME games. Any winning condition incorporating these constraints is a meager set.

However, not all winning conditions that are meager provide a winning strategy for Player 1. 0 has a winning strategy in an ME game $G$ just in case there is a sequence of moves $x$ such that for every finite prefix $x_n$ of $x$ ending in 1's turn, there is a finite prefix $y_n$ of plays by 0 such that $\mathcal{O}(x_n.y_n) \cap Win \neq \emptyset$.

8 For an introduction to LTL, see e.g. Lamport (1980).

And 1 has a winning strategy in $G$ just in case there is no such sequence $x$, or in other words just in case for every sequence of moves $x$ there is a finite prefix $x_n$ of $x$ ending in 1's turn, such that $\mathcal{O}(x_n) \cap Win = \emptyset$.

Consider the following abstract ME game. Suppose $V = \{a, b\}$ and suppose Player 0 loses if and only if at any point she plays $b$. That is, the winning set $Win$ is

$$Win = D \setminus Reach(\{(b, 0)\})$$

This is itself a rhetorically decomposition sensitive winning condition. Now, both $Win$ and $\pi_\omega(Win)$ are meager sets in their respective topologies. As the Banach Mazur theorem rightly states, Player 1 has a winning strategy in the BM game $(V^\omega, \pi_\omega(Win))$: play $b$ at some turn. However, she does not have a winning strategy in the ME game $(V_0 \cup V_1)^\omega$, $Win)$. That is because whatever she plays, Player 0 can always avoid playing $b$. In other words, the decomposition sensitivity of the ME games breaks down the applicability of the Banach Mazur theorem in ME games. Player 1 cannot 'play for' Player 0 now, which she can do in the BM game. A linguistic example of such a situation is a game $G$ where Janet from example (1) is player 0. Janet has a winning strategy in $G$ even though her winning condition is meager.

Conversely, consider a winning condition for 0 that depends on some finite number of contributions by 1. Call such goals *1-finite-decomposition sensitive*. An instance of such a winning condition would be Justin's. Recall that Justin's objective in (1) is to get Janet to commit as to whether she has been seeing Valentino or not. Symbolize this commitment by Janet's as $(c, 1) \in V_1$, as Janet is Player 1. Then Justin's winning condition is the union of open sets $\{\mathcal{O}(x.(c, 1)) \colon x \in (V_0 \cup V_1)^*\}$ and is co-meager. Nevertheless, his opponent Janet has a winning condition in such a game: never answer Justin's question directly. More generally,

**Proposition 6** *If an ME game $G$ has a 1-finite-decomposition sensitive winning condition, then there is no winning strategy in $G$ for* 0.

**Corollary 2** *There are ME games with 1-finite-decomposition sensitive winning conditions that are co-meager, but where* 0 *has no winning strategy.*

There are also 0-finite-decomposition sensitive winning conditions that depend only on finitely many moves of 0. These conditions are $\Sigma^0_1$ where 0 always has a winning strategy, as the BM theorem predicts. A final situation is the one of the prosecutor in example (2) has components that are decomposition sensitive, but the entire winning condition is not. The prosecutor's winning condition as described above is that Bronston must either commit to an answer or never answer $P$'s question. Given such a winning condition, there is a winning strategy for the prosecutor: keep asking the question until Bronston commits to an answer. In fact the entire game space is the winning condition for the prosecutor. Thus, we can infer:

**Proposition 7** *Decomposition sensitivity of winning conditions is not preserved under union.*

**Proof** Let $Win_1$ be a decomposition sensitive winning condition and let $Win_2 = D \setminus Win_1$. Clearly $Win_2$ is also decomposition sensitive. Indeed, because if the play is decomposed according to some play $u \in Win_1$ then Player 0 cannot win. However $Win_1 \cup Win_2 = D$ is clearly decomposition invariant. $\square$

We've now canvassed the whole spectrum of decomposition sensitive winning conditions in ME games. In general, decomposition sensitivity makes ME games more expressive and more complex than BM games, breaking the delicate link between topology and winning conditions given by the BM theorem.

Decomposition sensitivity also affects Borel complexity. If Player 0 has a winning strategy in a rhetorically decomposition sensitive ME game then the Borel complexity of $Win$ is at least $\Sigma^0_2$.

**Proposition 8** *Let $G = ((V_0 \cup V_1)^\omega, Win)$ be an ME game such that Win is rhetorically decomposition sensitive. If Player 0 has a winning strategy in G then the Borel complexiy of Win is at least $\Sigma^0_2$.*

**Proof** Let us mimic a winning play $u$ of Player 0. $u \in Win$ only if for every prefix $u_n$, $n$ odd, of $u$ that ends in a Player 1 move, $\mathcal{O}(u_n) \cap Win \neq \emptyset$. Thus whatever 1 plays to reach $u_n$, 0 can choose $v_n$ such that $u_{n+1} = u_n v_n$ is still a prefix of $u$. We may thus write $u$ as

$$u = u_0\mathcal{O}(u_1) \cap u_2\mathcal{O}(u_3) \cap \ldots$$

Hence $Win$, being at least a countable union of the above sequences, is at least $\Sigma^0_2$. $\square$

### 4.2 Winning Conditions: co-Büchi

We've already met the next kind of set in the Borel Hierarchy, $\Sigma^0_2$ sets, also known as co-Büchi sets. Suppose $C$ is a subset of $(V_0 \cup V_1)$ (the 'co-Büchi' set). Then

$$\textit{co-Büchi}(C) = \{x \in D \mid \inf(x) \subseteq C\}$$

where $inf(x) = \{a \in V \mid \forall i, \exists j > i, x(j) = a\}$ is defined to be the set of all the elements of $V$ which occur infinitely often in $x$.

In terms of LTL formulae, the co-Büchi condition may be viewed as follows. Let $C \subseteq V$ be the co-Büchi set. Then

$$\phi_{co-Buchi(C)} = \Diamond\Box \bigvee_{a \in C} p_a$$

We've also seen that rhetorically decomposition sensitive winning conditions, including consistency, responsiveness, coherence and NEC, are all at least $\Sigma^0_2$, if they have winning strategies. More particularly using proposition 8, if 0 must commit to at least some proposition to win—i.e., she has a $\Sigma^0_1$ objective—and must remain consistent, her winning condition is Co-Büchi. Classic examples of co-Büchi conditions are those with strings that eventually contain only elements of $C$ or eventually settle down in $C$. That is, the strings eventually get stuck in the safe set $C$. Example 5 is a motivating example for a conversation with a $\Sigma^0_2$ winning condition that also involves rhetorically decomposition sensitive conditions like responsiveness, coherence and consistency: Feynmann had to respond to his students' questions and in a coherent way lead them eventually to the topic that he wanted to discuss.

The winning condition that a conversation be finite is also a co-Büchi condition. A finite conversation is easily modeled in the ME framework; the initial segment in which the agreement is reached is then succeeded by an infinite sequence of "null" moves that keep the content of the last move. Indeed, there is a close connection between agreement winning conditions and finiteness. If the only goal of the exchange is to achieve a fixed point in which the dialogue stays within this information state forever after, the conversation should stop once the terms of the exchange and the agreement are common knowledge. Being rational agents, our players will stop once they acquire the mutual knowledge that that state has been achieved and that nothing will take them out of it.[9]

Bargaining agreements or agreements on some permanent exchange of goods, which could also be information are naturally $\Sigma^0_2$ conversations even in the absence of these constraints—e.g., . Any information seeking conversation in which 0 has the goal of acquiring agreement about some intellectual issue $\phi$, like a Socratic dialogue also has the structure of a $\Sigma^0_2$ winning condition. Walton (1984) calls these inquiry dialogues. Co-Büchi conditions distinguish between provisional and real agreement. In a provisional agreement, an agent

9 In principle, participants could continue acknowledging each other's acknowledgments *ad infinitem*. But such acknowledgments wouldn't serve any purpose. For a discussion see Asher & Venant (2015).

provisionally may acknowledge another's contribution and agree to a bargain but later take the acknowledgment and the agreement back. If *Win* of the conversational game consists only in reaching a provisional agreement, *Win* is clearly $\Sigma^0_1$, as it does not constrain what happens after the provisional agreement is reached. Real agreement is different. Once attained between two agents, the agents do not deviate from it in any further conversation; no conversational moves take them out of that state of agreement, as required for a co-Büchi condition. This is essentially also a decomposition sensitive condition.

In the absence of any constraints, however, a decomposition sensitive winning condition of agreement for 0 has no winning strategy for 0; 1 always has a non empty 1-play of disagreeing for any possible continuation by 0. Similarly, no conversational goal of extracting a binding oath from an opponent can succeed, unless additional constraints are imposed. For similar reasons to the lack of a winning strategy for agreement type winning conditions, finiteness winning conditions are also easily seen to have no winning strategy—1 can always prolong the game by talking when it is her turn.

Our general constraints of NEC, responsiveness and coherence, however, can make the agreement happen in agreement seeking conversations. The Jury becomes an "arbitratror", imposing agreement when the opponent 1 no longer has any counter arguments to rebut 0's arguments for a particular position or exchange; the lack of counter arguments makes 1's objections not credible, thus lowering 1's score eventually leading to 0's winning condition. Thus, with the Jury's constraints, 0 wins a $\Sigma^0_2$ goal iff 1 has eventually no more arguments against a certain proposition $\phi$, where $\phi$ may describe a bargain or topic of discussion.

Co-Büchi conditions also characterize goals in which 0 repeatedly attacks 1 eventually to reduce the opponent's score in the eyes of the Jury (Venant et al. 2014). (3) is such an example. Let LD be player 0 in an ME game. In the *voire-dire* transcript, 0 repeatedly returns to the question as to whether the defendant Tzeng was responsible for severing a nerve in a patient's hand; he seemed prepared to revisit the theme indefinitely until he exposes that the expert witness $D$, or 1 in this game, was covering up for a fault of the defendant. Repeatedly questioned, 1 replies each time in the play up to (3c,d) that Tzeng was not at fault. In the Jury's eyes, 0's questioning had little effect; the Jury's probabilities assigned to the types of 0 and 1 did not shift, and 0 was no closer to his winning condition in getting the court to agree with him that 1 was not an impartial witness. However, at (3c,d), 1 contradicts his previous testimony by agreeing to 0's loaded question, and his attempts to backtrack and correct his mistake are successfully attacked by 0 in (3h). At this point, 0 has achieved his goal.

We note that our model of the Jury needs refinement in that it does not take account of successful retractions in the face of inconsistency, and so we cannot really predict the counterattack at (3h). We plan to address this in future work.

### 4.3 Büchi conditions

The complementary condition of a Co-Büchi condition, the Büchi condition, is the equivalent of (infinite) iterated reachability, and is a condition that is not expressible on finite strings. Suppose $B$ is a subset of $(V_0 \cup V_1)$ (the 'Büchi' set). Then

$$\mathit{Büchi}(B) = \{x \in D \mid \inf(x) \cap B \neq \emptyset\}$$

is the set of all strings which contain infinitely many elements of $B$ or equivalently which visit $B$ infinitely often. A Büchi set is $\Pi^0_2$ in the Borel hierarchy. In conversations where player 0 has a Büchi winning condition, she will win if she always has a path to $B$ and revisits $B$ infinitely often. A Büchi condition is more sophisticated than a $\Pi^0_1$ condition, in which a player never leaves a set of states. 0 can play for a Büchi condition and allow 1 a reachability or $\Sigma^0_1$ condition on his play. In such a game, player 0 can continue to return to her chosen and preferred states infinitely often, reiterating a point or set of points that she wants to make (once again, a finite conjunction of Büchi conditions is also Büchi).

Some Büchi conditions can be expressed using LTL formulas, however, that are finitely satisfiable. For any $x \in V$, let the proposition $p_x$ denote the property of visiting or playing $x$. Let $B \subseteq V$ be the Büchi set of states. Then

$$\phi_{\mathit{Büchi}(B)} = \Box\Diamond \bigvee_{x \in B} p_x$$

If 0's winning condition means revisiting a set of states $B$ infinitely often, it must be for some other purpose other than agreements on exchanges of goods or information, for once lasting agreement is achieved, there is no point in revisiting that agreement. On the other hand, a Büchi condition can be effective in debate. Political debates like those evoked in example (6) exemplify a Büchi condition. Such a condition is more difficult to achieve if our rhetorical constraints are imposed on acceptable discourse sequences, because it means that any play by the opponent still must enable the player to have a rhetorically cooperative path to return to $B$. But a practiced debater can have such a strategy.

Let's now take a closer look at the analysis of one of our examples involving a Büchi winning condition, (4). Our excerpted example was a turning point in the Vice-Presidential debate. Quayle's goal as Player 0 was to continually revisit the theme that despite his youth he had the talent and experience of a good Vice-Presidential and Presidential candidate. In effect this is a $\Pi^0_2$ winning condition. Up to the exchange in 4, we can assume that Quayle had not made any disastrous moves, had remained consistent, responsive and replied to attacks and that the Jury's assignment to $\textsc{good}_0$ had not suffered that much. His play had produced an initial segment of strings in *Win*. That is, we assume that the play $\sigma$ up until (4) is such that $\|\sigma\|$ was above 0, though not significantly above.

Given this goal, it would seem 0 had a clear winning strategy. What went wrong? To describe the exchange in (4) in detail, we need as basic vocabulary for both $V_0$ and $V_1$: an attack move, $attack(x, y)$, meaning that the move $y$ attacks move $x$, descriptions of the content of basic moves $x\colon \phi$ and $y\colon \psi$, a commentary move, $comment(x, y)$ where a player expresses an opinion in $y$ about move $x$, and a question answering move (QAP). (4a) is a QAP move to a question about his Presidential qualifications. But the content of (4a) is ambiguous. Quayle might have just intended (4a)'s literal meaning—that he was equal in governmental experience to John Kennedy as a candidate for President. But he might also have intended, and probably did intend by mentioning a famous President, to have the audience and Jury draw a direct and positive comparison between himself and Kennedy with regards to the kind of President he might become in line with his winning condition. In response Bentsen plays $attack(4a, b).4b\colon \phi$, with $\phi$ contradicting the implicated direct comparison. This is what he should do; he should try to get the Jury to lower their estimation of Quayle's $\textsc{good}$ type. At this point Quayle should have counterattacked with another *attack* move, as NEC requires. But instead, Quayle plays a weak $comment(4b, c).4c\colon \psi$ in the subsequent turn; and then Bentsen plays another successful $attack(comment(4b, c), 4d).4d\colon \chi$, where $\chi$ says that it was Quayle brought up the comparison and thus opened himself up to attack—hence, $attack(4a, b)$ was perfectly fair. The Jury penalized Quayle severely for this failure to assume and defend the consequences of an implicature he most likely intended given his winning condition, setting $c_{attack(comment(4b,c),4d)} = 0$ and hence $P_{attack(comment(4b,c),4d)}(\textsc{good}_0) = 0$. Thus, at this point, $\|\sigma.4a - d\| = 0$, and Quayle could do nothing in the remainder of the debate to get $P_{4c}(\textsc{good}_0) > 0$. Given these assumptions, our model predicts that Quayle lost the debate with this one move.

## 4.4 Muller conditions

A Muller condition is defined as follows. Suppose we are given a set $\mathcal{F} = \{F_1, F_2, \ldots, F_n\}$ of subsets of $(V_0 \cup V_1)$ (the Muller sets). Then $\mathit{Muller}(\mathcal{F}) = \{x \in V^\omega \mid \mathit{inf}(x) \in \mathcal{F}\}$ is the set of all strings which eventually (after a finite point) get stuck in one of the Muller sets, $\mathit{Muller}(\mathcal{F})$.

A Muller winning condition is a boolean combination of Büchi and Co-Büchi conditions. In terms of temporal logic formulae this can be seen as follows. Let $\mathcal{F} = \{F_1, F_2, \ldots, F_n\}$ be the set of Muller sets where each $F_i$ is a subset of $V$. Then

$$\phi_{\mathit{Muller}(\mathcal{F})} = (\phi_{\mathit{co\text{-}Büchi}(F_1)} \vee \ldots \vee \phi_{\mathit{co\text{-}Büchi}(F_n)}) \wedge (\phi_{\mathit{co\text{-}Büchi}(F_1)} \Rightarrow$$

$$\bigwedge_{x \in F_1} \phi_{\mathit{Büchi}(\{x\})}) \wedge \ldots \wedge (\phi_{\mathit{co\text{-}Büchi}(F_n)} \Rightarrow \bigwedge_{x \in F_n} \phi_{\mathit{Büchi}(\{x\})})$$

Since Muller conditions extend Büchi conditions, Muller conditions are not compatible with the goal of exchanging goods. Nevertheless, there are real life conversations with Muller "winning" conditions with multiple states, in which the participants revisit the states indefinitely often. In fact conversations with Muller winning conditions are commonplace. For instance, examples (6) and (4) have both $\Pi^0_2$ and also $\Sigma^0_2$ components to their winning conditions, as the Jury requires that they obey rhetorical cooperativity, NEC and consistency. Once a $\Pi^0_2$ winning condition is combined with a $\Sigma^0_2$ requirement, the result is a Muller winning condition.

**Proposition 9** *If* 0 *must obey a rhetorically decomposition sensitive condition with a* $\Pi^0_2$ *objective, then her winning condition is Muller.*

There are also examples of conversations with Muller winning conditions. One involves a conversation between two partners who have lived for a long time together and who are quite old. After a certain point 0 always attempts to go through the same conversational moves, so that they can revisit the same memories, and touch on the same themes, laugh at the same jokes. 0 asks the same questions to get the same answers. To quote John Prine from the song *Far from Me, a question ain't really a question if you know the answer too.* In the song 1 plays along for a while, though she "waits a little too long," to laugh at the same, repeated jokes. In the end 1's goal is to break the cycle of repeated conversational moves by 0. Assuming that our conversational agents are rational, the goal of such a conversation is not information exchange or some sort of persuasion; it is something else like venting one's emotions albeit indirectly, reliving an experience, or conveying some other non literal message.

### 4.5 Relation to Gale-Stewart games

Another type of infinite games that have been extensively studied in the literature of Logic and Computer Science is called a Gale-Stewart game (GS game). A GS game is similar to a BM game in that the play of such a game is again an infinite sequence $x$ over a vocabulary $V$. However, whereas in a BM game the players take turns in playing finite non-empty *sequences* of letters from $V$, in a GS game, the players can play only single letters (from $V$) in each turn. In other words, the turn-structure of the play is built inherently into the dynamics of the game. Thus a GS game over a vocabulary $V$ is a tuple $G = (V, \mathit{Win})$ where $\mathit{Win}$, as before, is a subset of $V^\omega$. The notions of a strategy, winning strategy, determinacy etc are defined as a BM game. Every GS game $(V, \mathit{Win})$ where $\mathit{Win}$ is Borel is determined Martin (1975). However, the rigid turn structure of GS games precludes a characterization of the winner in terms of the topological properties of the winning set unlike in a BM game (thanks to the BM theorem). GS games in fact capture our ME games with decomposition sensitive winning conditions, while the BM games are isomorphic to the decompositive insensitive ME games.

The ME conversational games we proposed in this paper were developed by imposing a turn structure on BM games. Alternatively, we could have developed them as GS games (as the turn structure is built in) on the vocabulary $V^+ = V^* \setminus \{\epsilon\}$. Thus, the players take turns in playing elements from $V^+$ (which are finite non-empty sequences in $V$) forming an infinite sequence in $V^\omega$. A play can thus be viewed as a sequence in $((V_0^+ \cup V_1^+)^\omega \setminus (V_0^\omega \cup V_1^\omega)) = ((V_0 \cup V_1)^\omega \setminus (V_0^\omega \cup V_1^\omega)) = D$, where $V_i = V \times \{i\},\ i \in \{0,1\}$ as before. $\mathit{Win}$ once again, is a subset of $D$.

Exploiting the correspondence between GS games and decomposition sensitive ME games allows us to apply the Büchi-Landweber theorem, to infer the memory requirements of a winning strategy. To state this theorem we need a little background on what exactly is the memory of a strategy.

A strategy $s_0$ for Player 0, in a GS game $(V, \mathit{Win})$ is a function $s_0 : (V_0^+ V_1^+)^* \to V^+$ and a strategy $s_1$ for Player 1 is a function $s_1 : (V_0^+ V_1^+)^* V_0^+ \to V^+$. A strategy $s_i$ of player $i \in \{0,1\}$ is said to be *finite memory* if there exists a finite automaton with output $\mathcal{M}_i$ which dictates $s_i$. More formally let, $\mathcal{M}_i = (M_i, m_i^0, \delta_i, O_i)$ where $M_i$ is a finite set of states (the memory of $s_i$), $m_i^0 \in M_i$ is the initial memory, $\delta_i : V \times M \to M$ is the transition function and $O_i : V \times M \to V^+$ is the output function. Define the extended transition relation $\widehat{\delta}_i : V^+ \times M \to M$ from $\delta_i$ inductively as usual. That is, $\widehat{\delta}_i(v, m) = \delta_i(v, m)$ and $\widehat{\delta}_i(xv, m) = \delta_i(v, \widehat{\delta}_i(x, m))$. Then for every finite play $xv$ that ends in Player $(1-i)$s turn, $s_i(xv) = O_i(v, \widehat{\delta}_i(xv, m^0))$. $s_i$ is *memoryless* or *positional* if $M_i$ is a singleton. A positional strategy can be represented as a function $s_i : V \to V^+$. Now, the Büchi-Landweber theorem can be stated as follows

**Theorem 2 (Büchi & Landweber (1969))** *Let $G = (V, Win)$ be a Gale-Stewart game such that the Borel complexity of Win is at most a boolean combination of $\Sigma_2^0$ and $\Pi_2^0$ sets (that is, Win is at most Muller). Then one of the players always has a finite memory winning strategy.*

Coming back to conversations, this means that to win an ME game with a decomposition sensitive winning condition having a low Borel complexity, a player may require memory. However, if the winning condition is not 'too complicated', a finite amount of memory suffices. Here is an example illustrated with a particularly simple ME game with $V = \{a, b\}$:

(4) Consider an ME game $G = (V, Win)$ where 0 achieves *Win* iff she plays $a$ and $b$ alternatively in each of her turns. 1's moves do not matter.

0 has a winning strategy for $G$, but she has to remember what she did on her prior turn to achieve it. So she would need to have at least one cell of memory for a winning strategy. As another more linguistically sophisticated example, any winning condition that incorporates NR would require (2 cells of) turn memory.

As finite automata are one of the most tractable algorithmic objects, this suggests to us an ambitious but exciting future project: Given a debating situation between two (or more candidates) where the goals of the candidates can be represented as low-complexity Borel sets, predict the winner and design a winning strategy for her.

### 4.6 Finite satisfiability revisited

The examples of winning conditions that we've examined are all expressible as formulas of linear temporal logic (LTL). While LTL's semantics uses infinitary sequences of evaluation points for formulas, many of the formulas that capture intuitive winning conditions are also *finitely satisfiable*, satisfiable on finite sequences. For instance, a finite sequence $\sigma$ will verify a reachability goal $\Diamond\phi$ if its initial state does, and similarly for all the LTL definable conditions we have discussed. Even the LTL definable Muller conditions are Boolean combinations of LTL definable Büchi and Co-Büchi conditions, and so they also can be satisfied over finite sequences. In addition, since consistency or satisfiability and discourse coherence are defined for the finite formulas for $V$, finite prefixes of sequences in ME games can naturally also verify these properties, and our Jury winning condition declares a winner for any finite conversation.[10]

Thus, for example, in example 2, the prosecutor has achieved his $\Sigma^0_1$ of extracting at least a defeasible commitment from Bronston that he had no account and that both he and Bronston had coherent, responsive and consistent plays obeying NR and NEC. Given that the prosecutor's choice of winning condition was one the opinionated Jury found persuasive, it and the unbiased Jury would award the prosecutor a victory. However had the opinionated Jury required a stronger winning condition, on which the prosecutor had to extract a non-defeasible commitment from Bronston, it would have assigned Bronston the win. In example 3, LP's cornering D into an consistency yields LP a win, given our specification of the Jury.

The finite satisfiability of many winning conditions in the low levels of the Borel hierarchy leads us to ask about monotonicity or stability properties of these conditions. $\Sigma^0_1$ and $\Pi^0_1$ conditions are monotonic over finite sequences in the following sense: if $\phi$ is a $\Sigma^0_1$ condition and it is true in $\sigma$, then it is true in every extension of $\sigma$; conversely, if $\phi$ is a $\Pi^0_1$ condition and it is false in $\sigma$, then it is false in every extension of $\sigma$. However, $\Sigma^0_2$ and $\Pi^0_2$ goals are not monotonic in this sense without further conditions; a $\Sigma^0_2$ goal like an agreement may be reached in one finite sequence but then falsified in a finite continuation of that sequence, only to be reinstated in another continuation. This instability is reflected in the failure in our finitary semantics of certain LTL entailments. For instance a formula of the form $\Box\Diamond\phi$ can be satisfied in a finite sequence $x$ of length 1 (such that $\phi$ is true at the first index) but its LTL consequence $\bigcirc\Box\Diamond\phi$, where $\bigcirc$ is the next-time operator, is clearly false. This is another indication that $\Sigma^0_2$ and $\Pi^0_2$ are more complex than the sets at the first level of the Borel Hierarchy. The non-monotonic behavior of conditions at the second level of the Borel Hierarchy also reflects some of our linguistic observations—for instance, the difficulty of establishing lasting agreements over a finite fragment of conversation. Another example of this instability is this: NEC, the necessary condition on winning, may be falsified "unjustly", if 0 does not get the chance to reply to an attack. Perhaps for this reason, if the opponent launches an attack, most observers would require as fair that the attacked agent have the right of reply. Indeed all of our rhetorical constraints are unstable as can be seen from Proposition 9.

The non monotonic behavior of certain winning conditions over finite

10 Interestingly, conditions beyond $\Sigma^0_3$ in complexity do not have finitely satisfiable conditions; nor are they expressible using LTL formulae or even first order formulas of the language of linear orders McNaughton & Papert (1971).

prefixes allows us to clarify two ways to verify such conditions linking them to topological properties of the underlying space. There are two ways this sort of finite verification can fail: first, a condition may be verified in a finite prefix but verified only in some of its continuations and falsified on the others; secondly, a condition may never be verified on any finite prefix but only on infinite sequences. If all continuations of a finite prefix $s$ verify or fail to verify $\phi$, then the game is already decided at $s$. This happened in the debate from which example (4) is excerpted; after Bentsen's attack, our model predicts that Quayle's objectives were foiled, no matter how he continued. We note that the stable verification of $\phi$ in a finite string has a topological characterization. Consider all those infinite strings $s$ such that all of the finite prefixes of $s$ satisfy $\phi$. The set of all such strings $S$ is a model for $\phi$; if $S$ is open, then any game with $\phi$ as a winning condition is verified at a finite point, the point where $S$ starts to branch.

## 5 Conclusions and prospects

We have proposed a model for conversations in a strategic setting, ME games, building on BM games. We've provided a strong motivation for consistency rhetorical cooperativity, a key assumption of Asher & Lascarides (2013) even in the absence of other shared goals. And we've built these constraints into our conception of the Jury. We explored conversational goals which lie in the low levels of the Borel hierarchy with natural examples of conversations and provided a precise typology for them. We have shown how turns and turn involving winning conditions make ME games strictly more expressive than BM games and their winning strategies more complex. We also showed how intuitive constraints like consistency affect the complexity of winning conditions, moving many intuitive characterizations of simple winning conditions to $\Sigma_2^0$ or Muller in the Borel hierarchy. We've also shown some linguistic consequences of this typology with respect to stability or instability of evaluation.

We have just scratched the surface of a rich and interesting theory of the structure of conversations. There are many directions for further development. For instance, we have considered only very simple forms of our discourse constraints. Avoiding inconsistency in the more general case means avoiding a finite sequence of formulas in $V$ in accordance with the logic that yields a proof of an inconsistency. For first order logic, the consistency predicate is $\Pi_1^0$ over the language of Peano arithmetic, since provability is expressible as a $\Sigma_1^0$ predicate. It is not definable in S1S (or equivalently LTL), and so the complexity of consistency lies higher up in the Borel Hierarchy (or maybe even beyond that). We want to investigate more complex forms of consistency and other constraints with respect to the Borel Hierarchy and what this means for the existence and feasibility of winning strategies in ME games.

There are other directions for further research. One is to develop our model of the Jury further using techniques from mean-payoff games Zwick & Paterson (1995). Another is to investigate extensions of ME games to non zero sum settings by replacing simple winning conditions with utility functions, which would extend our analysis to many more conversations. Finally, we believe that integrating signaling games with ME games is exciting: each conversational move is in effect the result of a signaling game, and the ME game as a whole yields a sequence of iterated signaling games whose utilities are guided by the overall winning conditions.

## References


Asher, N. & A. Lascarides. 2003. *Logics of conversation*. Cambridge University Press.

Asher, N. & A. Lascarides. 2013. Strategic conversation. *Semantics and Pragmatics* Vol 6.2. 1–62.

Asher, N. & A. Venant. 2015. Ok or not ok? commitments, acknowledgments and corrections. In *Proceedings of semantics and linguistic theory (salt 25)*, Stanford.

Aumann, R. & S. Hart. 2003. Long cheap talk. *Econometrica* 71(6). 1619–1660.

Aumann, Robert J & Michael Maschler. 1995. *Repeated games with incomplete information*. MIT press.

Axelrod, Robert M. 2006. *The evolution of cooperation*. Basic books.

Benz, A. G. Jäger & R. van Rooij (eds.). 2005. *Game theory and pragmatics*. Palgrave Macmillan.

Büchi, J.R. & L.H. Landweber. 1969. Solving sequential conditions by finite-state strategies,. *Transactions of the American Mathematical Society* 138. 367–378.

Chang, C.C. & H. J. Keisler. 1973. *Model theory*. North Holland Publishing.

Crawford, V. & J. Sobel. 1982. Strategic information transmission. *Econometrica* 50(6). 1431–1451.

Farrell, J.1993. Meaning and credibility in cheap-talk games. *Games and Economic Behaviour* 5. 514–531.

Franke, M.2008. Meaning and inference in case of conflict. In Kata Balogh (ed.), *Proceedings of the 13th esslli student session*, 65–74.

Franke, M. 2009. *Signal to act: Game theory in pragmatics* ILLC dissertation series. Institute for Logic, Language and Computation. https://books.google.fr/books?id=dI6PZwEACAAJ.

Franke, M. T. de Jager & R. van Rooij. 2009. Relevance in cooperation and conflict. *Journal of Logic and Language* .

Glazer, J. & A. Rubinstein. 2001. Debates and decisions: On a rationale of argumentation rules. *Games and Economic Behavior* 36(2). 158–173.

Glazer, J. & A. Rubinstein. 2004. On optimal rules of persuasion. *Econometrica* 72(6). 119–123.

Grädel, E. 2008. Banach-Mazur games on graphs. In R. Hariharan, M. Mukund & V. Vinay (eds.), *Foundations of software technology and theoretical computer science (fsttcs)*, 364–382.

Grice, H. P. 1975. Logic and conversation. In P. Cole & J. L. Morgan (eds.), *Syntax and semantics volume 3: Speech acts*, 41–58. Academic Press.

Grosz, B. & C. Sidner. 1986. Attention, intentions and the structure of discourse. *Computational Linguistics* 12. 175–204.

Grosz, Barbara J. & Sarit Kraus. 1993. Collaborative plans for group activities. In *Proceedings of the thirteenth international joint conference on artificial intelligence*, 367–373. Los Altos, California: Morgan Kaufmann.

Kechris, A. 1995. *Classical descriptive set theory*. New York: Springer-Verlag.

Lamport, Leslie. 1980. Sometime is sometimes not never: On the temporal logic of programs. In *Proceedings of the 7th acm sigplan-sigact symposium on principles of programming languages*, 174–185. ACM.

Lewis, D.İ969. *Convention: A philosophical study*. Harvard University Press.

Mann, W. C. & S. A. Thompson. 1987. Rhetorical structure theory: A framework for the analysis of texts. *International Pragmatics Association Papers in Pragmatics* 1. 79–105.

Martin, D. A. 1975. Borel determinacy. *Annals of Mathematics* 102(2). 363–371.

Mauldin, R. (ed.). 1981. *The Scottish book. mathematics from the Scottish Café*. Birkhaüser.

McNaughton, Robert & Seymour Papert. 1971. Counter-free automata. In *Research monograph*, vol. 65, MIT Press.

Oxtoby, J. 1957. The Banach-Mazur game and Banach category theorem. *Contribution to the Theory of Games* 3. 159–163.

Parikh, P.İ991. Communication and strategic inference. *Linguistics and Philosophy* 14(5). 473–514.

Parikh, P. 2000. Communication, meaning and interpretation. *Linguistics and Philosophy* 25. 185–212.

Parikh, P.2̇001. *The use of language*. Stanford, California: CSLI Publications.

Rabin, M.İ990. Communication between rational agents. *Journal of Economic Theory* 51. 144–170.

Revalski, J. P. 2003-2004. The Banach-Mazur game: history and recent developments. Tech. rep. Institute of Mathematics and Informatics, Bulgarian

Academy of Sciences.
van Rooy, Robert. 2003. Being polite is a handicap: towards a game theoretical analysis of polite linguistic behavior. In *Tark*, 45–58.
Sacks, Harvey. 1992. *Lectures on conversation*. Oxford: Blackwell Publishers. Edited by Gail Jefferson. This is the published version of lecture notes from 1967–1972.
Solan, L.M. & P.M. Tiersma. 2005. *Speaking of crime: The language of criminal justice*. Chicago, IL: University of Chicago Press.
Spence, A. M. 1973. Job market signaling. *Journal of Economics* 87(3). 355–374.
Traum, D. & J. Allen. 1994. Discourse obligations in dialogue processing. In *Proceedings of the 32nd annual meeting of the association for computational linguistics (acl94)*, 1–8. Las Cruces, New Mexico.
van Rooij, R.2004. Signalling games select horn strategies. *Linguistics and Philosophy* 27. 493–527.
Venant, A., N. Asher & C. Dégremont. 2014. Credibility and its attacks. In *Proceedings of semdial 2014*, Edinburgh, Scotland: Semdial.
Walton, Douglas N. 1984. *Logical dialogue-games*. University Press of America, Lanham, Maryland.
Zwick, Uri & Michael S Paterson. 1995. The complexity of mean payoff games. In *Computing and combinatorics*, 1–10. Springer.